

STIC-Biotech/ChemLib

138619

From: Chan, Christina  
Sent: Tuesday, November 16, 2004 4:33 PM  
To: Walicka, Malgorzata  
Cc: STIC-Biotech/ChemLib  
Subject: RE:

Please rush. Thanks Chris

Chris Chan  
TC 1600 New Hire Training Coordinator and SPE 1644  
(571)-272-0841  
Remsen, 3E89

-----Original Message-----

From: Walicka, Malgorzata  
Sent: Tuesday, November 16, 2004 4:19 PM  
To: Chan, Christina  
Cc: STIC-Biotech/ChemLib  
Subject:

Christina,

Please authorize my RUSH search of SEQ ID NO:8 in 10/802,682. Thank you.

Malgorzata A. Walicka, Ph.D.  
Patent Examiner  
Art Unit 1652, Recombinant Enzymes  
USPTO, Remsen Building, Room 2C76  
400 Dulany St.  
Alexandria, VA 22313  
Mail Room 2C70  
Tel. (571)272-0944, fax (571)273-0944

RECEIVED  
NOV 16 2004  
STIC-BIOTECH/CHEM LIB DIVISION  
(STIC)

\*\*\*\*\*

STAFF USE ONLY

Searcher: P. Schneider  
Searcher Phone: 2- 2526  
Date Searcher Picked up: 11/18  
Date Completed: 11/18  
Searcher Prep/Rev. Time: 11  
Online Time: 5

\*\*\*\*\*

Type of Search

NA Sequence: # \_\_\_\_\_  
AA Sequence: # 1 \_\_\_\_\_  
Structure: # \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIS: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: Compuser  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2004, 04:20:35 ; Search time 159 Seconds  
(without alignments)  
1306.316 Million cell updates/sec

Title: US-10-802-682-8  
Perfect score: 3069  
Sequence: 1 MNPTLLRLTSAAVLLTAPA.....AEAIDSTSVGNALYFALPQ 579

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729239 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3069	100.0	579	2	AAW37876
2	3069	100.0	579	8	ADN10956
3	2765	90.1	580	8	AD134121
4	2611	85.1	579	2	AAW37873
5	2569.5	83.7	578	8	AD134118
6	2559.5	83.4	579	8	AD134120
7	2526.5	82.3	579	2	AAW37874
8	2497	81.4	578	2	AAW37875
9	1673	54.5	608	7	ADBE94131
10	585	19.1	754	3	AAW35987
11	552	18.0	685	7	ABO83287
12	551.5	18.0	738	7	AAW20192
13	550.5	17.9	738	2	AAW13993
14	490.5	16.0	742	2	AAW05235
15	412	13.4	803	7	ADDD24941
16	411	13.4	743	6	ABR42659
17	380.5	12.4	803	7	ADPF04027
18	375.5	12.2	818	6	ADA33921
19	367	12.0	740	2	AAW95019
20	341.5	11.1	826	6	ADA36228
21	316.5	10.3	543	8	ADH33692
22	302	9.8	602	7	ABO65919
23	286	9.3	688	7	ABO63296
24	216.5	7.1	706	7	ABO76984
25	175.5	5.7	381	6	ABU22217

26	161.5	5.3	243	7	ABO77068	AbO77068 Pseudomon
27	157.5	5.1	355	6	ABU41771	Abu41771 Protein e
28	157.5	5.1	381	6	ABU19930	Abu19930 Protein e
29	156.5	5.1	381	6	ABU21917	Abu21917 Protein e
30	154.5	5.0	380	6	ABU39856	Abu39856 Protein e
31	154.5	5.0	390	6	ABU40939	Abu40939 Protein e
32	154.5	5.0	399	7	ADFO7458	AdFO7458 Bacterial
33	147.5	4.8	443	3	ABG24430	AbG24430 Novel hum
34	145	4.7	119	5	ABO88379	ABO88379 Human ORF
35	140.5	4.6	391	6	ABW68770	ABW68770 Photornab
36	137	4.5	392	6	AAE36013	AAE36013 Moraxella
37	133.5	4.3	395	8	AD104851	AD104851 M. catarr
38	133	4.3	392	6	ABU35277	AbU35277 Protein e
39	133	4.3	392	6	ABU15441	AbU15441 Protein e
40	131	4.3	621	3	AAV59201	AAV59201 Bacillus
41	131	4.3	818	6	ABU20796	AbU20796 Protein e
42	130.5	4.3	3073	6	ABU21223	AbU21223 Protein e
43	129.5	4.2	386	6	ABU49222	AbU49222 Protein e
44	127.5	4.2	393	6	ABU50516	AbU50516 Protein e
45	127.5	4.2	948	4	ABG25909	AbG25909 Novel hum

## ALIGNMENTS

RESULT 1	AAW37876	standard; protein; 579 AA.
ID	AAW37876	
XX	AAW37876;	
AC	10-AUG-1998	(first entry)
DT	Alcohol and/or aldehyde dehydrogenase B amino acid sequence.	
XX	Alcohol and/or aldehyde dehydrogenase B enzyme; recombinant organism; aldehyde;	
XX	ketone; carboxylic acid; L-sorbose; D-sorbitol; 2-keo-L-gulonic acid;	
KW	L-ascorbic; inhibition.	
XX	Glucanobacter oxydans.	
OS	Key	Location/Qualifiers
XX	Peptide	1..23
FT	Protein	/note="signal peptide"
FT		24..579
FT		/note="mature protein"
XX	EP832974-A2.	
PN	01-APR-1998.	
XX	11-SEP-1997;	97EP-00115801.
XX	19-SEP-1996;	96EP-00115001.
XX	(HOFF) HOFFMANN LA ROCHE & CO AG F.	
PA	Asakura A, Hoshino T, Ojima S, Shinjoh W, Tomiyama N;	
PI	WPI, 1998-195228/18.	
DR	N-PSDB; AAV29054.	
XX	Recombinant Glucanobacter oxydans alcohol and/or aldehyde dehydrogenase	
PT	enzyme(s) - useful for converting substrate(s), e.g. L-sorbose or D-	
PT	sorbitol to 2-keo-L-gulonic acid.	
XX	Claim 1, Page 44-46; 59pp; English.	
PS	This is the amino acid sequence for the Glucanobacter oxydans alcohol	
XX	and/or aldehyde dehydrogenase B enzyme. The enzymes or recombinant	
CC	organisms can be used to convert suitable substrates to aldehydes,	
CC	ketones or carboxylic acids, especially to convert L-sorbose or D-	
CC	sorbitol to 2-keo-L-gulonic acid, which can be converted to L-ascorbic	

acid by standard procedures. The derivatives of AADH enzymes have desired substrate specificity, higher affinity to a substrate, lower affinity to an inhibitory compound, higher stability against temperature and/or pH and higher catalytic speed

Sequence 579 AA:

Query Match 100.0%; Score 3069; DB 2; Length 579;  
Best Local Similarity 100.0%; Pred. No. 4,3e-244;  
Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNPTTLRTSAVLLTAPAAFAQVPTITDELLANPAGEMINGRQENYRHSPLTQIT 60  
1 MNPTTLRTSAVLLTAPAAFAQVPTITDELLANPAGEMINGRQENYRHSPLTQIT 60  
QY 61 ADVNGQLQVWAKMEAGAVQVTPMIDGMYLANPBDVIOALDAQTGLIWEHRQLPA 120  
61 ADVNGQLQVWAKMEAGAVQVTPMIDGMYLANPBDVIOALDAQTGLIWEHRQLPA 120  
Db 61 ADVNGQLQVWAKMEAGAVQVTPMIDGMYLANPBDVIOALDAQTGLIWEHRQLPA 120  
QY 121 VATINAGDRKRGVALVGTSLYFSSMDNHLIALDMETGVVFPVERSGSDGLTSNTGP 180  
121 VATINAGDRKRGVALVGTSLYFSSMDNHLIALDMETGVVFPVERSGSDGLTSNTGP 180  
Db 121 VATINAGDRKRGVALVGTSLYFSSMDNHLIALDMETGVVFPVERSGSDGLTSNTGP 180  
QY 181 IVANGVIVAGSTCOYSPYGCFSIGHDSATGEBELRNHFIPQGEDEGTWGNDEARMT 240  
181 IVANGVIVAGSTCOYSPYGCFSIGHDSATGEBELRNHFIPQGEDEGTWGNDEARMT 240  
Db 181 IVANGVIVAGSTCOYSPYGCFSIGHDSATGEBELRNHFIPQGEDEGTWGNDEARMT 240  
QY 241 GWMQGITIDPVTNLVFGSTGVGPASSTQRTGCGTLYGNTFPAYRPDGEIWMHQTL 300  
241 GWMQGITIDPVTNLVFGSTGVGPASSTQRTGCGTLYGNTFPAYRPDGEIWMHQTL 300  
Db 241 GWMQGITIDPVTNLVFGSTGVGPASSTQRTGCGTLYGNTFPAYRPDGEIWMHQTL 300  
QY 301 PRDNDQECTFEMWVAVNDVQPSAEMGLAINPNAATGERRVLTPAECTGTWMSFDA 360  
301 PRDNDQECTFEMWVAVNDVQPSAEMGLAINPNAATGERRVLTPAECTGTWMSFDA 360  
Db 301 PRDNDQECTFEMWVAVNDVQPSAEMGLAINPNAATGERRVLTPAECTGTWMSFDA 360  
QY 361 SGEFLMARDNTYTNMISIDETGLVYNEDAVLKEIDVEYDVCPTLGGDMSAALNP 420  
361 SGEFLMARDNTYTNMISIDETGLVYNEDAVLKEIDVEYDVCPTLGGDMSAALNP 420  
Db 361 SGEFLMARDNTYTNMISIDETGLVYNEDAVLKEIDVEYDVCPTLGGDMSAALNP 420  
QY 421 TGIYFLPLNNAACVIMAVDOEFSAIDVYNTSATKLAPEENGRIDAIDISTRTLMSA 480  
421 TGIYFLPLNNAACVIMAVDOEFSAIDVYNTSATKLAPEENGRIDAIDISTRTLMSA 480  
Db 421 TGIYFLPLNNAACVIMAVDOEFSAIDVYNTSATKLAPEENGRIDAIDISTRTLMSA 480  
QY 481 ERPANYSPLVLTSTGIVENGCTDRYFRALSQETGELNQAALATVATGALISELDIQ 540  
481 ERPANYSPLVLTSTGIVENGCTDRYFRALSQETGELNQAALATVATGALISELDIQ 540  
Db 481 ERPANYSPLVLTSTGIVENGCTDRYFRALSQETGELNQAALATVATGALISELDIQ 540  
QY 541 YIAGAGLTGTGOLNAPLAFAIDSTSVGNAIVFALPQ 579  
541 YIAGAGLTGTGOLNAPLAFAIDSTSVGNAIVFALPQ 579  
Db 541 YIAGAGLTGTGOLNAPLAFAIDSTSVGNAIVFALPQ 579

RESULT 2  
ADN10956  
ID ADN10956 standard; protein; 579 AA.

XX AC ADN10956;

XX DT 01-JUL-2004 (first entry)

XX XX Gluconobacter oxydans Enzyme B, used in ascorbic acid production.

XX KW Enzyme B; ascorbic acid; vitamin C; L-gulonono-1,4-lactone; L-gulonic acid; L-galactono-1,4-lactone; L-galactonic acid.

XX OS Gluconobacter oxydans.

XX PN WO2004029267-A1.

XX PD 08-APR-2004.

XX PF 22-SEP-2003; 2003WO-EP010489.

PR 27-SEP-2002; 2002EP-00021602.

XX XX (STM) DSM IP ASSETS BV.

XX PI Hoshino T, Shinjoh M;

XX DR WPI; 2004-329889/30.

XX DR N-PSDB; ADN10955.

PT Producing L-ascorbic acid using enzyme B of Gluconobacter oxydans, from

PT substrates L-gulose, L-galactose, L-idose, and L-talose.

PS Claim 1; SEQ ID NO 2; 24pp; English.

The present sequence is the protein sequence of Enzyme B from Gluconobacter oxydans strain DSM 4025. Enzyme B has a molecular weight of about 60,000 Da by SDS-PAGE, substrate specificity for primary and secondary alcohols and aldehydes, is stable in the pH range 6-9 with optimal activity at about pH 8.0, and is inhibited by Cu<sup>2+</sup>, Zn<sup>2+</sup>, Mn<sup>2+</sup>, Fe<sup>2+</sup> and Fe<sup>3+</sup>. The present invention provides the use of this enzyme in a process for producing L-ascorbic acid from L-gulose, L-galactose, L-idose or L-talose, or from L-gulonono-1,4-lactone, L-gulonic acid, L-galactono-1,4-lactone, L-galactonic acid, L-idono-1,4-lactone, L-idonic acid, L-talono-1,4-lactone and L-talonic acid. Enzyme B is also used in a process for the production of L-gulonono-1,4-lactone or L-gulonic acid from L-gulose, and L-galactono-1,4-lactone or L-galactonic acid from L-galactose. The processes involve contacting the enzyme with the respective substrate and isolating the product from the reaction mixture. The process is conducted for 1-20 hours at pH 1-9 (preferably pH 2-8) and 13-45 (preferably 18-42) degrees C. Production of L-gulonono-1,4-lactone/L-gulonic acid from L-gulose, vitamin C from L-gulonono-1,4-lactone/L-gulonic acid, L-galactono-1,4-lactone/L-galactonic acid from L-galactose, and vitamin C from L-galactono-1,4-lactone/L-galactonic acid by recombinant *Escherichia coli* JM109 carrying the Enzyme B gene is described in examples from the invention.

XX XX Sequence 579 AA:

Query Match 100.0%; Score 3069; DB 8; Length 579;  
Best Local Similarity 100.0%; Pred. No. 4,3e-244;  
Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNPTTLRTSAVLLTAPAAFAQVPTITDELLANPAGEMINGRQENYRHSPLTQIT 60  
1 MNPTTLRTSAVLLTAPAAFAQVPTITDELLANPAGEMINGRQENYRHSPLTQIT 60  
Db 1 MNPTTLRTSAVLLTAPAAFAQVPTITDELLANPAGEMINGRQENYRHSPLTQIT 60  
QY 61 ADVNGQLQVWAKMEAGAVQVTPMIDGMYLANPBDVIOALDAQTGLIWEHRQLPA 120  
61 ADVNGQLQVWAKMEAGAVQVTPMIDGMYLANPBDVIOALDAQTGLIWEHRQLPA 120  
Db 61 ADVNGQLQVWAKMEAGAVQVTPMIDGMYLANPBDVIOALDAQTGLIWEHRQLPA 120  
QY 121 VATINAGDRKRGVALVGTSLYFSSMDNHLIALDMETGVVFPVERSGSDGLTSNTGP 180  
121 VATINAGDRKRGVALVGTSLYFSSMDNHLIALDMETGVVFPVERSGSDGLTSNTGP 180  
Db 121 VATINAGDRKRGVALVGTSLYFSSMDNHLIALDMETGVVFPVERSGSDGLTSNTGP 180  
QY 181 IVANGVIVAGSTCOYSPYGCFSIGHDSATGEBELRNHFIPQGEDEGTWGNDEARMT 240  
181 IVANGVIVAGSTCOYSPYGCFSIGHDSATGEBELRNHFIPQGEDEGTWGNDEARMT 240  
Db 181 IVANGVIVAGSTCOYSPYGCFSIGHDSATGEBELRNHFIPQGEDEGTWGNDEARMT 240  
QY 241 GWMQGITIDPVTNLVFGSTGVGPASSTQRTGCGTLYGNTFPAYRPDGEIWMHQTL 300  
241 GWMQGITIDPVTNLVFGSTGVGPASSTQRTGCGTLYGNTFPAYRPDGEIWMHQTL 300  
Db 241 GWMQGITIDPVTNLVFGSTGVGPASSTQRTGCGTLYGNTFPAYRPDGEIWMHQTL 300  
QY 301 PRDNDQECTFEMWVAVNDVQPSAEMGLAINPNAATGERRVLTPAECTGTWMSFDA 360  
301 PRDNDQECTFEMWVAVNDVQPSAEMGLAINPNAATGERRVLTPAECTGTWMSFDA 360  
Db 301 PRDNDQECTFEMWVAVNDVQPSAEMGLAINPNAATGERRVLTPAECTGTWMSFDA 360  
QY 361 SGEFLMARDNTYTNMISIDETGLVYNEDAVLKEIDVEYDVCPTLGGDMSAALNP 420  
361 SGEFLMARDNTYTNMISIDETGLVYNEDAVLKEIDVEYDVCPTLGGDMSAALNP 420  
Db 361 SGEFLMARDNTYTNMISIDETGLVYNEDAVLKEIDVEYDVCPTLGGDMSAALNP 420  
QY 421 TGIYFLPLNNAACVIMAVDOEFSAIDVYNTSATKLAPEENGRIDAIDISTRTLMSA 480  
421 TGIYFLPLNNAACVIMAVDOEFSAIDVYNTSATKLAPEENGRIDAIDISTRTLMSA 480  
Db 421 TGIYFLPLNNAACVIMAVDOEFSAIDVYNTSATKLAPEENGRIDAIDISTRTLMSA 480

QY 481 ERPAANYSFVLSTAGGVVFNCGTDRYFRALSOETGETLMQARLATVATGQALSYELDVQ 540  
 DB 481 ERPAANYSFVLSTAGGVVFNCGTDRYFRALSOETGETLMQARLATVATGQALSYELDVQ 540  
 QY 541 YTAIGAGGLTYGTQLNAPLAERAIIDSTSVGNAYVFALPQ 579  
 DB 541 YTAIGAGGLTYGTQLNAPLAERAIIDSTSVGNAYVFALPQ 579

RESULT 3  
 AD134121  
 ID AD134121 standard; protein; 580 AA.  
 AC AD134121;  
 XX 15-APR-2004 (first entry)  
 DE Ketogulononicigenium sp. sorbitol dehydrogenase (SDH) 3 protein.  
 XX  
 KM Sorbitol dehydrogenase; SDH; cytochrome c; molecular biology;  
 KM bacteriology; industrial fermentation; L-sorbose; 2-keto-L-gulonic acid;  
 KM 2KUG; enzyme.  
 OS Ketogulononicigenium.  
 XX  
 XX Key Location/Qualifiers  
 FT Peptide 1..23  
 FT /label= Signal\_peptide  
 FT Protein 24..580  
 FT /note= "Mature SDH protein"  
 XX  
 PN US2003228672-A1.  
 PD 11-DEC-2003.  
 XX  
 PF 06-JUN-2002; 2002US-00162713.  
 XX  
 PR 06-JUN-2002; 2002US-00162713.  
 XX  
 PA (ARCH) ARCHER-DANIELS MIDLAND CO.  
 XX  
 PI Choi E, D'elja J, Kim H, Kim M, Lee JK, Pan J, Stoddard SF,  
 PI Yum D;  
 XX  
 DR MPI; 2004-052025/05.  
 DR N-PSDB; AD134117, AD134124.  
 XX  
 FT New proteins of sorbitol dehydrogenases and cytochrome c of the strains  
 FT Ketogulononicigenium spp., useful in molecular biology, bacteriology and  
 FT industrial fermentation, e.g. for producing L-sorbose and 2-keto-L-  
 FT gulonic acid.  
 XX  
 PS Claim 79; SEQ ID NO 8; 68pp; English.  
 XX  
 CC The invention relates to the identification and isolation of nucleic acid  
 CC sequences and proteins of sorbitol dehydrogenases (SDH) and cytochrome c  
 CC of the strains, Ketogulononicigenium sp. The proteins and nucleic acid  
 CC molecules are useful in the fields of molecular biology, bacteriology and  
 CC industrial fermentation specifically for producing L-sorbose and 2-keto-L-  
 CC gulonic acid (2KUG). The present sequence is Ketogulononicigenium sp. SDH  
 CC protein.  
 XX  
 SQ Sequence 580 AA;

Query Match 90.1%; Score 2765; DB 8; Length 580;  
 Best Local Similarity 87.6%; Pred. No. 5.2e-219;  
 Matches 507; Conservative 38; Mismatches 34; Indels 0; Gaps 0;

QY 1 MNPTLLTSAVLLITPAAPAOVTPTTDELLANPPAGEWINRGROENTRHSPLQIT 60  
 DB 1 MRPTLLTSAVLLITPAAPAOVTPTTDELLANPPAGEWINRGROENTRHSPLQIT 60

QY 61 ADNVGLOLVARAMEAGAVQTPMIDHGVNVLNAPGDVIOALDAQTGLIMEHRRLP 120  
 DB 61 TINVGLOLVARAMEAGAVQTPMIDHGVNVLNAPGDVIOALDAQTGLIMEHRRLP 120  
 QY 121 VATLNAGDRKRGVALLYGTSLYFSSWMDNHLALDMETQGVVDVERGSGEDDLTNTTGP 180  
 DB 121 VASINQGDPRKRGVALLYGTSLYFSSWMDNHLALDMETQGVVDVERGSGEDDLTNTTGP 180  
 QY 181 IVANGVIVAGSTCOQSPYGCFTSGHDSATGEELMKNHFIPOGEGEDETWGNDFEARMYT 240  
 DB 181 IVANGVIVAGSTCOQSPYGCFTSGHDSATGEELMKNHFIPOGEGEDETWGNDFEARMYT 240  
 QY 241 GVMGQITVDPVNTNLVFGYSTGVGPASEORGTGGTLGINTRFAPRPTGTEIVRHQTL 300  
 DB 241 GVMGQITVDPVNTNLVFGYSTGVGPASEORGTGGTLGINTRFAPRPTGTEIVRHQTL 300  
 QY 301 PRDNMDQECTFEMMVANVDVQPSAMEGLRALNPAAATGERAVLTGAPCKGTWMSFDA 360  
 DB 301 PRDNMDQECTFEMMVANVDVQPSAMEGLRALNPAAATGERAVLTGAPCKGTWMSFDA 360  
 QY 361 SGEFLMARDNTYTNMIASIDETGLTVNEDAVLXELDYEDVCPFFIGSRDMSAALPD 420  
 DB 361 TGEFLMARDNTYTNMIASIDETGLTVNEDSVLTQDLDDYDICTFFLGGSDMPSAALPD 420  
 QY 421 TGIYFPLNNACVDIMAVDQEFSAIDVYNTSATKILAPGFENMGRIIDAIDISTGRTLWSA 480  
 DB 421 SGIFYFPLNNACVDIMAVDQEFSAIDVYNTSATKILAPGFENMGRIIDAIDISTGRTLWSA 480  
 QY 481 ERPAANYSFVLSTAGGVVFNCGTDRYFRALSOETGETLMQARLATVATGQALSYELDVQ 540  
 DB 481 ERPAANYSFVLSTAGGVVFNCGTDRYFRALSOETGETLMQARLATVATGQALSYELDVQ 540  
 QY 541 YTAIGAGGLTYGTQLNAPLAERAIIDSTSVGNAYVFALPQ 579  
 DB 541 YTAIGAGGLTYGTQLNAPLAERAIIDSTSVGNAYVFALPQ 579

RESULT 4  
 AAM37873  
 ID AAM37873 standard; protein; 579 AA.  
 XX  
 AC AAM37873;  
 XX  
 DT 10-AUG-1998 (first entry)  
 XX  
 DE Alcohol and/or aldehyde dehydrogenase A amino acid sequence.  
 XX  
 KM Alcohol/aldehyde dehydrogenase A enzyme; recombinant organism; aldehyde;  
 KM ketone; carboxylic acid; L-sorbose; D-sorbitol; 2-keto-L-gulonic acid;  
 KM L-ascorbic; inhibition.  
 XX  
 OS Gluconobacter oxydans.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..23  
 FT /note= "signal peptide"  
 FT Protein 24..579  
 FT /note= "mature protein"  
 XX  
 PN EP832974-A2.  
 XX  
 PD 01-APR-1998.  
 XX  
 PF 11-SEP-1997; 97EP-00115801.  
 XX  
 PR 19-SEP-1996; 96EP-00115001.  
 XX  
 PA (HOF) HOFMANN LA ROCHE & CO AG F.  
 XX  
 PI Aakura A, Hoshino T, Ojima S, Shinjoh M, Tomiyama N;  
 XX  
 DR MPI; 1998-195228/18.  
 DR N-PSDB; AAV29051.

XX Recombinant Gluconobacter oxydans alcohol and/or aldehyde dehydrogenase  
PT enzyme(s) - useful for converting substrate(s), e.g. L-sorbose or D-  
PT sorbitol to 2-keto-L-gulononic acid.

PS Claim 1; Page 35-37; 59pp; English.

XX This is the amino acid sequence for the Gluconobacter oxydans alcohol  
CC and/or aldehyde dehydrogenase A enzyme. The enzymes or recombinant  
CC organisms can be used to convert suitable substrates to aldehydes,  
CC ketones or carboxylic acids, especially to convert L-sorbose or D-  
CC sorbitol to 2-keto-L-gulononic acid, which can be converted to L-ascorbic  
CC acid by standard procedures. The derivatives of ADH enzymes have desired  
CC substrate specificity, higher affinity to a substrate, lower affinity to  
CC an inhibitory compound, higher stability against temperature and/or pH  
CC and higher catalytic speed

XX Sequence 579 AA;

Query Match 85.1%; Score 2611; DB 2; Length 579;  
Best Local Similarity 82.1%; Pred. No. 2,78-206;  
Matches 476; Conservative 51; Mismatches 51; Indels 2; Gaps 2;

QY 1 MNPTTLRTSAVLLTAPAAFAQVPTITDELLANPAGEMWNGRQENYRHSPLTQIT 60  
DB 1 MKPTSLMASAGALALAAAPFAQVPTITDELLANPAGEMWISYGRQENYRHSPLTQIT 60  
QY 61 ADVNGQLQVWARGMEAGAVQVTPMIDGVWYLANPGDVIOALDACTGDLIWEHRQLPA 120  
DB 61 TENNGQLQVWARGMEAGAVQVTPMIDGVWYLANPGDVIOALDACTGDLIWEHRQLPA 120  
QY 121 VATNAGQDRKRGVALGTSLYFSSWMDNHLIALDMETGOVVPVDRSGEDGLTSNTGP 180  
DB 121 IATNLSGEPTRGALGTSLYFSSWMDNHLIALDMETGOVVPVDRSGEDGLTSNTGP 180  
QY 121 IATNLSGEPTRGALGTSLYFSSWMDNHLIALDMETGOVVPVDRSGEDGLTSNTGP 180  
DB 121 IATNLSGEPTRGALGTSLYFSSWMDNHLIALDMETGOVVPVDRSGEDGLTSNTGP 180  
QY 181 IVANGVIVAGSTCOYSPYGCIFSGHDSATGSELMRNHFIPOGEGDETGNDFEARWMT 240  
DB 180 IVANGVIVAGSTCOYSPYGCIFSGHDSATGSELMRNHFIPOGEGDETGNDFEARWMT 240  
QY 241 GWMQGITVDPVTNLVFGYSTGVGASSTQRTPGTGILYGNTRFAVPDGTGEIVMHHQTL 300  
DB 240 GWMQGITVDPVTNLVFGYSTGVGASSTQRTPGTGILYGNTRFAVPDGTGEIVMHHQTL 300  
QY 301 PRDWMDOCTEFEMWVAVNDVOPSAEMEGRLAINPNAATGERRVLTGAPCKTGTWMSFDA 360  
DB 300 PRDWMDOCTEFEMWVAVNDVOPSAEMEGRLAINPNAATGERRVLTGAPCKTGTWMSFDA 360  
QY 361 SGEFLMARDNYTMMASIDETGLVTNEDAVLKELDVETDVCPTPLGGRDMSAALNDP 420  
DB 360 SGEFLMARDNYTMMASIDETGLVTNEDAVLKELDVETDVCPTPLGGRDMSAALNDP 420  
QY 421 TGIFFLPNNACVYIMAVDOEFSAIDVNTSATKLAPEENGRIDAIDISTGRTLMSA 480  
DB 420 SGIFFLPNNACVYIMAVDOEFSAIDVNTSATKLAPEENGRIDAIDISTGRTLMSA 480  
QY 481 ERPAANTSPVISTAGVFNNGTDTRYFRALSOETGETLMQARLATATGCAIYEILDQV 540  
DB 480 ERPAANTSPVISTAGVFNNGTDTRYFRALSOETGETLMQARLATATGCAIYEILDQV 540  
QY 541 YIAIAGAGLTGTOLNAPLA-EAIDSTSVGNAIVFVLPQ 579  
DB 540 YIAIAGAGLTGTOLNAPLA-EAIDSTSVGNAIVFVLPQ 579

RESULT 5  
AD134118  
ID AD134118 standard; protein; 578 AA.

XX AD134118;  
XX 15-APR-2004 (first entry)  
XX Ketogulononicigenium sp. sorbitol dehydrogenase (SDH) 1 protein.  
DE

XX Sorbitol dehydrogenase; SDH; cytochrome c; molecular biology;  
KW bacteriology; industrial fermentation; L-sorbose; 2-keto-L-gulononic acid;  
KW 2KLG; enzyme.

OS Ketogulononicigenium.

FT Key Location/Qualifiers  
FT Peptide 1..23  
FT Protein 24..578  
FT /note= "Mature SDH protein"

EN US2003228672-A1.

PD 11-DEC-2003.

PF 06-JUN-2002; 2002US-00162713.

PR 06-JUN-2002; 2002US-00162713.

PA (ARCH) ARCHER-DANIELS-MIDLAND CO.

PI Choi E, D'elia J, Kim H, Kim M, Lee JK, Pan J, Stoddard SF;  
PI Yum D;

DR WPI: 2004-052025/05.  
DR N-PSDB; AD134114, AD134122.

PT New proteins of sorbitol dehydrogenases and cytochrome c of the strains  
PT Ketogulononicigenium spp., useful in molecular biology, bacteriology and  
PT industrial fermentation, e.g. for producing L-sorbose and 2-keto-L-  
XX gulonic acid.

PS Claim 67; SEQ ID NO 5; 68pp; English.

XX The invention relates to the identification and isolation of nucleic acid  
CC sequences and proteins of sorbitol dehydrogenases (SDH) and cytochrome c  
CC of the strains, Ketogulononicigenium sp. The proteins and nucleic acid  
CC molecules are useful in the fields of molecular biology, bacteriology and  
CC industrial fermentation specifically for producing L-sorbose and 2-keto-L-  
CC gulonic acid (2KLG). The present sequence is Ketogulononicigenium sp. SDH  
CC protein.

XX Sequence 578 AA;

Query Match 83.7%; Score 2569.5; DB 8; Length 578;  
Best Local Similarity 80.1%; Pred. No. 78-203;  
Matches 464; Conservative 59; Mismatches 55; Indels 1; Gaps 1;

QY 1 MNPTTLRTSAVLLTAPAAFAQVPTITDELLANPAGEMWNGRQENYRHSPLTQIT 60  
DB 1 MKNSTLLASVAAVAFVAPFADVTPTDELLANPAGEMWISYGRQENYRHSPLTQIT 60  
QY 61 ADVNGQLQVWARGMEAGAVQVTPMIDGVWYLANPGDVIOALDACTGDLIWEHRQLPA 120  
DB 61 PDNNGQLQVWARGMEAGAVQVTPMIDGVWYLANPGDVIOALDACTGDLIWEHRQLPA 120  
QY 121 VATNAGQDRKRGVALGTSLYFSSWMDNHLIALDMETGOVVPVDRSGEDGLTSNTGP 180  
DB 121 TSTLSSLDKRGALVYGTNYFVSWDHWVALDAASGVYFVDVDRSGEDERV-SNNSGP 179  
QY 181 IVANGVIVAGSTCOYSPYGCIFSGHDSATGSELMRNHFIPOGEGDETGNDFEARWMT 240  
DB 180 IVANGVIVAGSTCOYSPYGCIFSGHDSATGSELMRNHFIPOGEGDETGNDFEARWMT 240  
QY 241 GWMQGITVDPVTNLVFGYSTGVGASSTQRTPGTGILYGNTRFAVPDGTGEIVMHHQTL 300  
DB 240 GWMQGITVDPVTNLVFGYSTGVGASSTQRTPGTGILYGNTRFAVPDGTGEIVMHHQTL 300  
QY 301 PRDWMDOCTEFEMWVAVNDVOPSAEMEGRLAINPNAATGERRVLTGAPCKTGTWMSFDA 360  
DB 300 PRDWMDOCTEFEMWVAVNDVOPSAEMEGRLAINPNAATGERRVLTGAPCKTGTWMSFDA 360

QY 361 SGEFLMARDNTYNNMTASIDETGLATVNEBDALVELKEDVEYDVCPTFLGGSDMSAALNDP 420  
 Db 360 TGEFLMARDNTYNNMTASIDETGLATVNEBDALVELKEDVEYDVCPTFLGGSDMSAALNDP 419  
 QY 421 TGIYFLPLNNACYDIMAADVEFSALDVNTSATAKLAPGFENMGRIADIDISTGRTLMSA 480  
 Db 420 SGIFYFLPLNNACADLAADVEFSALDVNTSATYLLAPEKENMGRIDAIDISTGRTLMSV 479  
 QY 481 ERPAANYSPLSTAGGVFENGSDRYFRALSEETGETLWQTRLATVAGSGLSYELDGVQ 540  
 Db 480 ERLASNSPVLSTAGGVFENGSDRYFRALSEETGETLWQTRLATVAGSGLSYELDGVQ 539  
 QY 541 YIAIGAGGLTYGTOLNAPLAELAIDSTSVGNAIYVFALPQ 579  
 Db 540 YVIAAGGNTYGTINLNSNIGATIDSTISGNAYVVFALPQ 578

RESULT 6  
 ADI34120  
 ID ADI34120 standard; protein; 579 AA.  
 AC ADI34120;  
 DT 15-APR-2004 (first entry)  
 DE Ketoglucigenium sp. sorbitol dehydrogenase (SDH) 2 protein.  
 XX Sorbitol dehydrogenase; SDH; cytochrome c; molecular biology;  
 KM bacteriology; industrial fermentation; L-sorbitol; 2-keto-L-gulononic acid;  
 XX 2KLG; enzyme.  
 OS Ketoglucigenium.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..23  
 FT Protein /label= signal\_peptide  
 FT 24..579  
 FT /note= "Mature SDH protein"  
 FT  
 XX US2003228672-A1.  
 PN 11-DEC-2003.  
 XX  
 PF 06-JUN-2002; 2002US-00162713.  
 XX  
 PR 06-JUN-2002; 2002US-00162713.  
 XX  
 PA (ARCH) ARCHER-DANIELS MIDLAND CO.  
 XX  
 PI Choi B, D'elia J, Kim H, Kim M, Lee JK, Pan J, Stoddard SF;  
 PI Yun D;  
 DR WPI; 2004-052025/05.  
 DR N-PSDB; ADI34116, ADI34123.  
 XX  
 PT New proteins of sorbitol dehydrogenases and cytochrome c of the strains  
 PT Ketoglucigenium spp., useful in molecular biology, bacteriology and  
 PT industrial fermentation, e.g. for producing L-sorbitol and 2-keto-L-  
 PT gulonic acid.  
 XX  
 PS Claim 75; SEQ ID NO 7; 68bp; English.  
 XX  
 CC The invention relates to the identification and isolation of nucleic acid  
 CC sequences and proteins of sorbitol dehydrogenases (SDH) and cytochrome c  
 CC of the strains, Ketoglucigenium sp. The proteins and nucleic acid  
 CC molecules are useful in the fields of molecular biology, bacteriology and  
 CC industrial fermentation specifically for producing L-sorbitol and 2-keto-L-  
 CC -gulonic acid (2KLG). The present sequence is Ketoglucigenium sp. SDH  
 CC protein.  
 XX  
 SO Sequence 579 AA;

Query Match 83.4%; Score 2559.5; DB 8; Length 579;  
 Best Local Similarity 80.5%; Pred. No. 4.7e-202;  
 Matches 466; Conservative 52; Mismatches 60; Indels 1; Gaps 1;

QY 1 MNPFTLSTSAVLLITPAAPQVPTPTDELLANPPAGEMINNGRQENRHSPLQIT 60  
 Db 1 MKTSEFLPAGVAAALASVGTIALADVTPTDELLANPPAGEMISYGRQENRHSPLQIT 60  
 QY 61 ADNVGQLQVVARGMEAGAVQVTPMIDGVVYLANPQGVIALDAQCTGLIWEHRRLQPA 120  
 Db 61 PENVGQLQVVARGMEAGAVQVTPMIDGVVYLANPQGVIALDAQCTGLIWEHRRLQPA 120  
 QY 121 VATLNAQGRDRRGVALYGTSLYFSSMDNHLIALDMETGVQVFDYERSGEDGLTSNTTGP 180  
 Db 121 VATLNSFGEPIRGIALYGTNYVYFVSMNHLIALDAATQVTFDYDRGGED-NTVSNSSGP 179  
 QY 181 IVANGVIVAGSTCOYSPYGCPISGHDSATGEBLNRNHFIPQGEGETWGNDFEARMYT 240  
 Db 180 IVANGVIVAGSTCOYSPGCEVSGHDATGEBLNRNHFIPQGEGETWGNDFEARMYT 239  
 QY 241 GVMGQITVDPVTNLVYFGSTGVGPASETQRTPGTLYGNTTRPAVRPDGEIYWRHQT 300  
 Db 240 GVMGQITVDPVTNLVYFGSSAVGPASETQRTGGTGMGTNTTRPAVRPDGEIYWRHQT 299  
 QY 301 PRDNDQECTFEEMVAVNDVQPSAEMGLRALNPPAAGERRVLTGAFCCKGTWMSPDAA 360  
 Db 300 PRDNDQECTFEEMVAVNDVQPSADMDLKSINPPAAGERRVLTGVCKGTWMSQDAE 359  
 QY 361 SGEFLMARDNTYNNMTASIDETGLATVNEBDALVELKEDVEYDVCPTFLGGSDMSAALNDP 420  
 Db 360 TGEFLMARDNTYNNMTASIDETGLATVNEBDALVELKEDVEYDVCPTFLGGSDMSAALNDP 419  
 QY 421 TGIYFLPLNNACYDIMAADVEFSALDVNTSATAKLAPGFENMGRIADIDISTGRTLMSA 480  
 Db 420 SGIFYFLPLNNACADLAADVEFSALDVNTSATYLLAPEKENMGRIDAIDISTGRTLMSV 479  
 QY 481 ERPAANYSPLSTAGGVFENGSDRYFRALSEETGETLWQTRLATVAGSGLSYELDGVQ 540  
 Db 480 ERLASNSPVLSTAGGVFENGSDRYFRALSEETGETLWQTRLATVAGSGLSYELDGVQ 539  
 QY 541 YIAIGAGGLTYGTOLNAPLAELAIDSTSVGNAIYVFALPQ 579  
 Db 540 YVIAAGGNTYGTINLNSNIGATIDSTISGNAYVVFALPQ 578

RESULT 7  
 AAM37874  
 ID AAM37874 standard; protein; 579 AA.  
 AC AAM37874;  
 DT 10-AUG-1998 (first entry)  
 DE Alcohol and/or aldehyde dehydrogenase A' amino acid sequence.  
 XX  
 XX Alcohol/aldehyde dehydrogenase A' enzyme; recombinant organism; aldehyde;  
 KM ketone; carboxylic acid; L-sorbitol; D-sorbitol; 2-keto-L-gulononic acid;  
 KM L-ascorbic; inhibition.  
 XX  
 OS Gluconobacter oxydans.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..23  
 FT Protein /note= "signal peptide"  
 FT 24..579  
 FT /note= "mature protein"  
 PN EP832974-A2.  
 XX  
 PD 01-APR-1998.  
 XX  
 PF 11-SEP-1997; 97EP-00115801.  
 XX

PR 19-SEP-1996; 96EP-00115001.  
 XX (HOFF) HOFFMANN LA ROCHE & CO AG F.  
 XX  
 PI Asakura A, Hoshino T, Ojima S, Shinjoh M, Tomiyama N;  
 XX WPI; 1998-195228/18.  
 DR N-PSDB; AAV29052.  
 XX  
 PT Recombinant Gluconobacter oxydans alcohol and/or aldehyde dehydrogenase  
 PT enzyme(s) - useful for converting substrate(s), e.g. L-sorbose or D-  
 PT sorbitol to 2-keto-L-gulonic acid.  
 XX  
 PS Claim 1; Page 38-40; 59pp; English.  
 XX  
 CC This is the amino acid sequence for the Gluconobacter oxydans alcohol  
 CC and/or aldehyde dehydrogenase A' enzyme. The enzymes or recombinant  
 CC organisms can be used to convert suitable substrates to aldehydes,  
 CC ketones or carboxylic acids, especially to convert L-sorbose or D-  
 CC sorbitol to 2-keto-L-gulonic acid, which can be converted to L-ascorbic  
 CC acid by standard procedures. The derivatives of AADH enzymes have desired  
 CC substrate specificity, higher affinity to a substrate, lower affinity to  
 CC an inhibitory compound, higher stability against temperature and/or pH  
 CC and higher catalytic speed  
 XX  
 SQ Sequence 579 AA:  
 Query Match 82.3%; Score 2526.5; DB 2; Length 579;  
 Best Local Similarity 79.8%; Pred. No. 2.5e-199;  
 Matches 462; Conservative 56; Mismatches 60; Indels 1; Gaps 1;

RESULT 8

AA037875  
 ID AA037875 standard; protein; 578 AA.  
 XX  
 AC AA037875;  
 XX  
 DT 10-AUG-1998 (first entry)  
 XX  
 DE Alcohol and/or aldehyde dehydrogenase A' amino acid sequence.  
 XX  
 KW Alcohol/aldehyde dehydrogenase A' enzyme; recombinant organism;  
 KW aldehyde; ketone; carboxylic acid; L-sorbose; D-sorbitol;  
 KW 2-keto-L-gulonic acid; L-ascorbic; inhibition.  
 XX  
 OS Gluconobacter oxydans.  
 XX  
 XX Key Location/Qualifiers  
 FH Peptide 1..23  
 FT /note= "signal peptide"  
 FT Protein 24..578  
 FT /note= "mature protein"  
 XX  
 FN EP832974-A2.  
 XX  
 PD 01-APR-1998.  
 XX  
 PF 11-SEP-1997; 97EP-00115801.  
 XX  
 PR 19-SEP-1996; 96EP-00115001.  
 XX  
 PA (HOFF) HOFFMANN LA ROCHE & CO AG F.  
 XX  
 PI Asakura A, Hoshino T, Ojima S, Shinjoh M, Tomiyama N;  
 XX WPI; 1998-195228/18.  
 DR N-PSDB; AAV29053.  
 XX  
 PT Recombinant Gluconobacter oxydans alcohol and/or aldehyde dehydrogenase  
 PT enzyme(s) - useful for converting substrate(s), e.g. L-sorbose or D-  
 PT sorbitol to 2-keto-L-gulonic acid.  
 XX  
 PS Claim 1; Page 41-43; 59pp; English.  
 XX  
 CC This is the amino acid sequence for the Gluconobacter oxydans alcohol  
 CC and/or aldehyde dehydrogenase A' enzyme. The enzymes or recombinant  
 CC organisms can be used to convert suitable substrates to aldehydes,  
 CC ketones or carboxylic acids, especially to convert L-sorbose or D-  
 CC sorbitol to 2-keto-L-gulonic acid, which can be converted to L-ascorbic  
 CC acid by standard procedures. The derivatives of AADH enzymes have desired  
 CC substrate specificity, higher affinity to a substrate, lower affinity to  
 CC an inhibitory compound, higher stability against temperature and/or pH  
 CC and higher catalytic speed  
 XX  
 SQ Sequence 578 AA:  
 Query Match 81.4%; Score 2497; DB 2; Length 578;  
 Best Local Similarity 79.3%; Pred. No. 6.7e-197;  
 Matches 459; Conservative 57; Mismatches 61; Indels 2; Gaps 2;



QY 241 GWMGQITYPVNLVIFYSSTGVGPASSETQRTGPGTLTGNTTPAVNPDTGELIWMHQTL 300  
 DB 239 GWMGQITYPVNLVIFYSSTGVGPASSETQRTGPGTLTGNTTPAVNPDTGELIWMHQTL 298  
 QY 301 PRDMDCECFEEMVAVNDVQPSAEMEGRLAINDNATGERRVLGAPCKTGTMSPDA 360  
 DB 299 PRDMDCECFEEMVAVNDVQPSAEMEGRLAINDNATGERRVLGAPCKTGTMSPDA 358  
 QY 361 SGEFLMARDNTYNTMISIDETGLVTVNEBDVAVKELDVAVCPPTFGSDMSAALNP 420  
 DB 359 TGEFLMARDNTYNTMISIDETGLVTVNEBDVAVKELDVAVCPPTFGSDMSAALNP 418  
 QY 421 TGIYFLPLNNACTDINAVDOFSLADYNTSATKLAAPGEMNGRIDALISTGRTLSA 480  
 DB 419 TGIYFLPLNNACTDINAVDOFSLADYNTSATKLAAPGEMNGRIDALISTGRTLSA 478  
 QY 481 ERPAANTPYLSTAGVAVFNGGTRDYFRALSOETGETLMQARLATVATGQAISEYLDGV 540  
 DB 479 ERPAANTPYLSTAGVAVFNGGTRDYFRALSOETGETLMQARLATVATGQAISEYLDGV 538  
 QY 541 YVIAAGGLTYGTQNLNAPLAELDSTSVGNALIVFALPQ 579  
 DB 539 YVIAAGGLTYGTQNLNAPLAELDSTSVGNALIVFALPQ 577

## RESULT 9

AD94131  
 ID ADE94131 standard; protein; 608 AA.

AC ADE94131;

DT 12-FEB-2004 (first entry)

DE Alcohol/aldehyde dehydrogenase, SEQ ID 1.

KM Enzyme; alcohol/aldehyde dehydrogenase; pyrrolo quinoline quinone;

KM L-sorbose; L-sorbose; 2-keto-L-gulonic acid.

OS Pseudoglucobacter saccharoketogenes; IF014464.

XX JF2003159079-A.

XX 03-JUN-2003.

XX 29-NOV-2001; 2001JP-00364508.

XX 29-NOV-2001; 2001JP-00364508.

XX (FUJI ) FUJISAWA PHARM CO LTD.

XX WPI; 2003-818681/77.

XX N-PSDB; ADE94132.

PT Novel alcohol/aldehyde dehydrogenase protein catalyzing oxidation of

PT hydroxymethyl group of compound to aldehyde group and aldehyde group of

PT compound to carboxyl group, useful for manufacturing 2-keto-L-gulonic

PS Claim 3; SEQ ID NO 1; 48pp; Japanese.

XX The present invention relates to an alcohol/aldehyde dehydrogenase (1;

XX ADE94131), which catalyzes oxidation of hydroxymethyl group of a compound

XX to an aldehyde group and the aldehyde group of a compound to a carboxyl

XX group. (1) does not contain hemoferrum and rare earth elements. (1) has

XX an optimum pH of 4.5-5.5, isoelectric point of 4.1 +/- 0.3, pyrrolo

XX quinoline quinone as a prosthetic group and K<sub>m</sub> value of 40 mM for

XX sorbose. (1) is useful for manufacturing a compound having a carboxyl

XX group by contacting the compound having hydroxymethyl group or an

XX aldehyde group with (1) where the compound having a hydroxymethyl group

XX or an aldehyde group is L-sorbose or L-sorbose and the compound having

XX a carboxyl group is 2-keto-L-gulonic acid.

SQ Sequence 608 AA;

Query Match 54.5%; Score 1673; DB 7; Length 608;

Best Local Similarity 52.5%; Pred. No. 6.9e-129;

Matches 315; Conservative 94; Mismatches 167; Indels 24; Gaps 6;

QY 2 NPTTLIRSAVLLTAAAPRQ-----VPIITBELLANPAGWI 42  
 DB 9 NVVGLATSTALIASGSP-AFAQHDANAAPSSKAGQAIENFQVPTADDLGKXNPANP 67  
 QY 43 NYGRQENVRHSPILQITADNVGQQLVWARGMEAGAVQTPMIDGMYLANPQDVQA 102  
 DB 68 IIRGVQGMGYSPLDQIKNDVNGDQLVWMSRTMEGSGAIAVNGVIFLGNTNDVQA 127  
 QY 103 LDAQTGDLIMEHRRLPVAAT-LNAQGRKGVALYGTSLYSSMDNHLIALDMETGVV 161  
 DB 128 IDGKTGSLIMEYRRRLPASKEFINSIGAARSIALLFGDKYFVSDNFFVALDAKATGELA 187  
 QY 162 FVERGSGEDGLTNTGTPTPIVANGVIVAGSTCOXSPYGCFTSGHDSATGEELMRNHFIPQ 221  
 DB 188 WETNRGQYEBGVANSSGPVIVDGVILAGSTCOXSGFECYVTGIDAESGEELMRNHFIPR 247  
 QY 222 PGEEDGETWGN-DFEARMWTGVWQITYPVTVNLYFGSTGVGPASSETQRTGPGTLVGT 280  
 DB 248 PGEEDDITWGPAPYENRMWTGAMQITYPDELIVYSGTGVGPASSETQRTGPGTLVGT 307  
 QY 281 NTRFAVRPDTGELIYVRHQTLPDNDNDCEFTFEMVAVNDVQPSAEMEGRLAINDNATG 340  
 DB 308 NTRFAVRPDTGELIYVRHQTLPDNDNDCEFTFEMVAVNDVQPSAEMEGRLAINDNATG 367  
 QY 341 -SRVLTPGAPCKTGWMSFPAASGEFLMARDNTYNTMISIDETGLVTVNEBDVAVKELDVE 399  
 DB 368 TEKVLTPGAPCKTGWMSFPAASGEFLMARDNTYNTMISIDETGLVTVNEBDVAVKELDVE 427  
 QY 400 YDVCPTFLGGRDMSAALNPDTGTYFLPLNNACTDINAVDOFSLADYNTSATKLAAPG 459  
 DB 428 YVYCTFLGGRDMSAALNPDTGTYFLPLNNACTDINAVDOFSLADYNTSATKLAAPG 487  
 QY 460 FENMGRIDALIDSTRTLSAERPAANTPYLSTAGVAVFNGGTRDYFRALSOETGETLM 519  
 DB 488 KTMGRVADILATSETKMSYETRALYDPVLTITGDLVFGGIDRDRALDAESGKRW 547  
 QY 520 QARLATVATGQAISEYLDGVQYIALGAGGLTYGTQNLNAPLAELDSTSVGNALIVFALPQ 579  
 DB 548 STRLPGAIVSGYTTYSIDGRQYVAVVSGG-SUGGPTFGPTTDPDVASGANGIYVFLPQE 606

## RESULT 10

AD94131  
 ID AAB35987 standard; protein; 754 AA.

AC AAB35987;

DT 01-MAR-2001 (first entry)

DE Sorbitol dehydrogenase subunit 1 amino acid sequence.

KM Sorbitol dehydrogenase; SDH; Gluconobacter oxydans; D-sorbitol;

KM L-sorbose production; 2-keto-L-gulonic acid.

XX Gluconobacter oxydans.

XX WO200065066-A1.

XX 02-NOV-2000.

XX 23-APR-1999; 99MO-1B000736.

XX 23-APR-1999; 99MO-1B000736.

XX (CHOI/) CHOI E.

XX (RHEB/) RHEB S.

XX (LEER/) LEER E.

XX Choi E, Rhee S, Lee E;  
 XX WPI: 2000-687351/67.  
 DR N-PSDB; AAC83153.  
 XX Novel membrane-bound sorbitol dehydrogenase nucleic acid molecule  
 PT isolated from Gluconobacter suboxydans useful for the fermentative  
 PT production of 2-keto-L-gulonic acid and L-sorbose from D-sorbitol.  
 XX  
 PS Claim 1; Fig 8; 96pp; English.  
 CC This invention relates to an isolated membrane-bound sorbitol  
 CC dehydrogenase (SDH) from Gluconobacter oxydans. The invention includes  
 CC polynucleotides (AAC83153 - AAC83154) encoding the 3 subunits of SDH  
 CC (AAB35987 - AAB35989). Also included in the invention are two  
 CC polynucleotide sequences AAC83156 and AAC83157 which contain the subunit  
 CC coding regions. The sorbitol dehydrogenase polynucleotide sequences are  
 CC useful for producing L-sorbose from D-sorbitol and for increasing the  
 CC production of 2-keto-L-gulonic acid by transforming a host cell,  
 CC especially Gluconobacter with the DNA and selecting the transformed host  
 CC cell. Note: An additional 66 DNA sequences (fragments of the SDH subunit  
 CC genes) encoding fragments of SDH are specifically claimed, however these  
 CC sequences are not given separately in the specification but are included  
 CC in sequences AAC83156 and AAC83157  
 CC  
 XX Sequence 754 AA;  
 SQ  
 Query Match 19.1%; Score 585; DB 3; Length 754;  
 Best Local Similarity 28.2%; Pred. No. 5,5e-39;  
 Matches 172; Conservative 94; Mismatches 238; Indels 106; Gaps 21;  
 6 LLRTSAAN-LLTAPAAFAQ--VTPTDELLANPPAGMINTGRNENYRHSPLTQTD 62  
 16 LLGCAALAFCAISPALADTGTALTN--ADQHPEDWMSYKRTTSEORISPLDITD 72  
 63 NVGQLQVW-----ARGMEGAVOYTPMIDGVYLANPGVIALDAQGLDWEHR 116  
 73 NASNLKLAHMYDDITNRQGE-----TPLIDGVMTATTWSMKALDAATGKLMSYD 127  
 117 QLRP-VATLNAQDRKRGVALYGTSLYFSSWDHLLALDMENQGVVFDY----- 164  
 128 KVPENIADRGCCDVTNRGAAYMNGKYFETFDRLIALAKTGKLAWSVYVTPKRAQLH 187  
 165 ERSGGEDGLSNTTGPVANG-VIVAGSTCOYSPYCGFISGHDSATGEELMNFIPGP 223  
 188 QRSYTDG-----APRIAKGXIIIGNGAEPARG-FVYAVAEKGKDMRFETYPND 240  
 224 EEDG-----ETWQNDPEARKMTG--VWQITYPYTNLVYVYTGCVGPASE 267  
 241 NKEDGAASDVLMSKAVPTWKGGAWKQGGGCTVDSLIDPVTDLVLYGVNGSPMY 300  
 268 TQGTGPGTLYGNTFRFAVPDGTGETVWRHQLPRDNMQDECTFEWVNAVDPQSAWE 327  
 301 KFSSEGKNNLFLGSIVAINPDGKYVMHFQETPMQMDYTSVQIMA--LDMPVNGEN- 357  
 328 GLRAINPNATGERRVLTGAPCKTGTWVSFDAAGEFLMARDTYNMTIASIDE-TGLVT 386  
 358 -----RHVLVHAP-KNGFFYIIDAKTGKISGKPYTENMANGLDPTVGRPN 403  
 387 VNEDAVLXELDYVDVDEFTLGRDWSAALNDGTGYFLPLNNACVYDIMAVDQESALD 446  
 404 YNPDALMTLNGKWPYGI-PDLSGHNPAVAAYSPQTKLVIPAOQVF--VTDPOGKXK 460  
 447 VYNTSATAKLPGFENNGRIDAIDISTGR-----TLMSAERPAANY-- 488  
 461 AHHDSNMLGL-----DNKRIGLDDNDPQHKADKAQFLKDLKGMIVAMPQKQAAFTVD 515  
 489 -----PVLTAGVYFNGGTRYPALASQETGETLMQARLATATGALISYELDGYX 541  
 516 HKSPMNGGLATAGGVLFQGLANGEPFAVDATTKDLPFPAOSATIAAPVTYTAGKXY 575  
 542 IA--IGAGGL 549

DB 576 VAVEWGMGI 585  
 RESULT 11  
 ID ABO83287 standard; protein; 685 AA.  
 XX ABO83287;  
 AC ABO83287;  
 XX  
 DT 29-JUL-2004 (first entry)  
 XX  
 DE Pseudomonas aeruginosa polypeptide #15462.  
 XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.  
 XX Pseudomonas aeruginosa.  
 XX US6551795-B1.  
 XX 22-APR-2003.  
 PD 18-FEB-1999; 99US-00252991.  
 PF 18-FEB-1998; 98US-0074788P.  
 PR 27-JUL-1998; 98US-0094190P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
 PI WPI: 2003-615309/58.  
 DR N-PSDB; ABD16858.  
 PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
 PT pathological conditions resulting from bacterial infection.  
 PS Disclosure: SEQ ID NO 32033; 455pp; English.  
 XX  
 CC The invention relates to Pseudomonas aeruginosa polypeptides and the  
 CC polynucleotides encoding them. The sequences are useful in diagnosis and  
 CC therapy of pathological conditions, as molecular targets for diagnostics,  
 CC prophylaxis and treatment of pathological conditions resulting from a  
 CC bacterial infection, for evaluating a compound, such as a polypeptide,  
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
 CC effective antibacterial targets, as targets for antibacterial drugs,  
 CC including anti-P. aeruginosa drugs, as templates for recombinant  
 CC production of P. aeruginosa-derived peptides or polypeptides, as target  
 CC components for diagnosis and/or treatment of P. aeruginosa-caused  
 CC infection, and in detection of P. aeruginosa sequences or other sequences  
 CC of Pseudomonas species using biochip technology. Sequences ABO67826-  
 CC ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from USPTO at  
 CC seqdata.uspto.gov/sequence.html  
 CC  
 XX Sequence 685 AA;  
 SQ  
 Query Match 18.0%; Score 552; DB 7; Length 685;  
 Best Local Similarity 27.2%; Pred. No. 2,5e-36;  
 Matches 180; Conservative 110; Mismatches 220; Indels 142; Gaps 28;  
 3 PTLIRISA-----AVLLTPAAFAQVTPITDELLAN--PPAGMINTGRNENYRHS 55  
 71 FAGLRPSLHCLAPAAVAGSAGALAK--DYTWEDIANDDKTTGDVLYQGMGTGAQWSP 128  
 56 LTOITADNVGQLQVWMA-----RGMGAAGVQVTPMIDGVYLANPGDVIALDAQT 107  
 129 LKQVNAADVFLCTANYSRGDEKRGQESQAI-----VSDGVITVYASYSRI-PALDAKT 183  
 108 GDLIWEHRRL-----PAVATLNAQDRKRGVALYGTSLYFSSWDHLLALDMETGVVF 162

Db 184 GKRLMTYNHRLPDIRECCDVN-----RGAALYGDKVFFGLDASVALNKNTKYVW 237  
 QY 163 DVERSGEDGLTSNTGPIVANG-----VIVAGST-CQYSPYGCFTSGHDSATGELMEN 216  
 Db 238 KKRPRADGAGVYMTGATLYVDKGTGVLLIHSSSGSEFVGRFLR-RPPDGEIWMR 296  
 QY 217 HFLPQPEEG-----DETGNDEPA-----RMWTG---WGQITYPVTN 253  
 Db 297 PFV--EGHMGRLNGKSTVTGVDKAPSPDDRNSPTGKVESWGHGGAPQASFAETN 354  
 QY 254 LVFVGTGVPASETORGTGPG-----TLVGTNTPRAVPRDGEIWMHQLPBDNMWQ 307  
 Db 355 TLIVGAGNPPPMNTWMTAKGNPHDYSLY-TSGQGVDPSSGEVWTFYQHPNDAMP 413  
 QY 308 ECTFEMVAVNVDPQSAEMEGELAINPNAATGERVLTGAPCKTGTWMSFDAASGE----- 363  
 Db 414 SGNNELVL-----FDYKADGKIVKATAHADNNGFFVYVDRNGKLONA 457  
 QY 364 -----FLWADNTYTMIMASID-ETGLVTVNED--AVLKEIDVEY-----DVCEFTLGR 410  
 Db 458 PPVVDNITWA-----SHIDIKTRPVREGRQRPPLPEPQKIGKAVEVSPFLGK 508  
 QY 411 DWSSAALNPDTGIFLPLNNAACYDIMAVDQESALDVYNTSATKLAAPG-----ENK 463  
 Db 509 NMNPMAYSDTGLFYVANH-----WKEDVTEBVSYTKSAYLGMGFRIRKMYDDHY 561  
 QY 464 GRIDAIISTGRITMSAERPANYSPLYSTAGGVENGSTDRYFRALSOETGETLMQAL 523  
 Db 562 GSIRAMPVSGKVWMEKEHLPLMAGVLTATAGNLVFTGTDGYPKAFDAKSGKELMKFOR 621  
 QY 524 ATVATGQAISEYELDVQY--IAIGAGGLT--YG--TQLNAPLAEAIDSTSVGNALYVA 576  
 Db 622 GSGIVSPITWEDQGEYLGVTGYGGAIVPMGDMADLTPVAQ-----GSGTWVRK 674  
 QY 577 LP 578  
 Db 675 LP 676  
 RESULT 12  
 AAR20192 standard; protein; 738 AA.  
 XX AAR20192;  
 AC AAR20192;  
 XX 27-AUG-2003 (revised)  
 DT 16-APR-1992 (first entry)  
 XX ADH complex protein (mol.wt. 72.000).  
 XX Alcohol dehydrogenase; acetic acid; fermentation.  
 XX Acetobacter sp.  
 OS Acetobacter sp.  
 XX JF03266988-A.  
 XX 27-NOV-1991.  
 XX 26-MAR-1990; 90JP-00073440.  
 XX 26-FEB-1990; 90JP-00042391.  
 XX (NAKA-) NAKANO SUTEN KK.  
 XX Tamaki T, Takemura H, Tayama K, Fukaya M, Okumura H, Kawamura Y,  
 XX WPI: 1991-288462/40.  
 XX N-PSDB; AAQ20383.  
 XX Gene for membrane-bound alcohol dehydrogenase complex - obcd. from  
 PT Acetobacter alioacetigenes, used for prodn. of enzyme for converting  
 PT alcohol to acid.  
 XX

PS Disclosure; Fig 3 (1-3); 21pp; Japanese.  
 XX Acetobacter transformed with the sequence encoding this protein can  
 CC enhance the efficiency of acetic acid fermentation. The ADH complex can  
 CC be easily extracted from the bacteria and purified and it can be used for  
 CC the determination of an alcohol. See also AAQ20383-84, and -86-88.  
 CC (updated on 27-AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 738 AA;  
 Query Match 18.0%; Score 551.5; DB 2; Length 738;  
 Best Local Similarity 27.0%; Pred. No. 3,1e-36;  
 Matches 164; Conservative 91; Mismatches 253; Indels 99; Gaps 17;  
 QY 5 TLRTSAVLLLTAPAEPAQVTP--ITDELL--ANPPAGEWVINGENQENYRHSPLTQI 59  
 Db 15 TLTAGTICAAALISGATWASADGGATGEAIIHADDPGNMWTYGRYSQRYSPLDQI 74  
 QY 60 TADNVGOQLVW-----ARGMEAGAVQVTEMHGVMYLANPBDVIALDAQGDLWE 113  
 Db 75 NRSNVGNLKLAWYLDLDTNRQEG---TPVLDGVNVAITTNMSWMAVDAATGKLMS 129  
 QY 114 HRRQLPA-VATLNAQGRKRGVALYGTSLYFSSWMDNHLIADMETGQVFPV----- 164  
 Db 130 YDPRVPGNLADKCCDDTVNRGAAYWNGKVYFGTPEGRLIALDAKTGKLVMSVNTIPPEAE 189  
 QY 165 ---ERSGEDGLTSNTGPIVANG-VIVAGSTCQYSPYGCFTSGHDSATGELWNRHFI 220  
 Db 190 LKQKSYTVDG-----APRIAKGRVITGNQSGSEFGANG-FVSAPDAETGVDRFFFTVP 242  
 QY 221 QPGEED-----ETWGNDFEARMMWG-----TWGQITYPVTNLVFPY 258  
 Db 243 NPKMPPDAASVLMNKAIVQTS-----PTGANTRGGGGGTWDSLVYPPVADLVLG 295  
 QY 259 STGVGPASETORGTGPGTLVGTNTRFAVRPDTGEIWMHQLPBDNMDOCTFEMMVAVN 318  
 Db 296 VNGSGPMVYKYSEEGKGNLFLGSIYALKPTEGYVMHFOETPMQWFTSDQIMTIDL 355  
 QY 319 DVQPAEWEGLRAINPNAATGERRVLTGAPCKTGTWMSFDAASGEFLWARTNTNMLAS 378  
 Db 356 PI-----NGSTRVIVHARKNGFTYIIDKKTGEFISGKNYVYVNMASG 398  
 QY 379 ID-ETGLVTVNEDAVLKEIDVEYDVCPTFLGRDSSAALNPDTGIFLPLNNAACY----- 433  
 Db 399 LDKTGRPIYNDALYTLTGKEMWGIPEGDLGHNFAANA-FPKGLVYIIPAOVPELYTN 458  
 QY 434 -----DIMAVDQESALDVYNTSATKLAAPGEMKRIIDALDISTGRTLSAERPA 484  
 Db 459 QVGFPTPHPSWNLGLDMNKVGIPID-SPEAKQAFVKDLKGIIVAWDPQKQAEAWVDHKG 517  
 QY 485 ANYSPVLSTAGGVENGSTDRYFRALSOETGETLMQALAVATGQAISEYELDVQYTA- 543  
 Db 518 PMNGGILATGDDLFOGLANGEFHAYDATNGSDLFHFAADGGIAPRTYIYLANGQYAV 577  
 QY 544 -IGAGGL 549  
 Db 578 EVWGMI 584  
 RESULT 13  
 AAR13993 standard; protein; 738 AA.  
 XX AAR13993;  
 AC AAR13993;  
 XX 27-AUG-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 09-DEC-1991 (first entry)  
 XX A.alioacetigenes membrane-bound ADH 72kD sub-unit.  
 DE alcohol dehydrogenase complex; carboxylic acid production.  
 XX

OS Acetobacter sp.  
 XX EP448969-A.  
 PN  
 PD 02-OCT-1991.  
 XX  
 PF 26-FEB-1991; 91EP-00102793.  
 XX  
 PR 26-FEB-1990; 90JP-00042391.  
 PR 26-MAR-1990; 90JP-00073440.  
 XX  
 PA (NAKA-) NAKANO VINEGAR CO LTD.  
 XX  
 PI Tamaki T, Takemura H, Tayama K, Fukaya M, Okumura H, Kawamura Y;  
 DR WPI; 1991-289462/40.  
 DR N-PSDB; AAQ13580.  
 XX  
 PT Gene for membrane-bound alcohol dehydrogenase complex - obtd. from  
 PT Acetobacter alcohogenes, used for prodn. of enzyme for converting  
 PT alcohol to acid.  
 PS  
 PS Disclosure; Fig 3; 36pp; English.  
 XX  
 CC Total DNA was prepared from A.alcohogenes MH-24, digested with PstI and  
 CC SmaI and ligated to PstI-SmaI cleaved pUC18. The ligation mixture was  
 CC used to transform E.coli JM109. Probes were designed based on the N-  
 CC terminal amino acid sequence of the ADH complex isolated from A.  
 CC alcohogenes (see AAQ13582-Q13584). The gene encoding the 72kD ADH  
 CC subunit was isolated and sequenced. The directly sequenced N-terminal  
 CC region of the purified 72kD sub-unit corresponds to the sequence  
 CC beginning at residue 36 of the deduced sequence. This suggests that the  
 CC first 35 N-terminal amino acids form a leader peptide involved in  
 CC secretion of the mature 72kD protein. The deduced amino acid sequence has  
 CC 77 percent homology with the same enzyme from A. aceti K6033. See also  
 CC AAQ13581. (Updated on 25-MAR-2003 to correct PA field.) (updated on 27-  
 CC AUG-2003 to correct OS field.)  
 CC  
 XX  
 SQ Sequence 738 AA;  
 Query Match 17.9%; Score 550.5; DB 2; Length 738;  
 Best Local Similarity 27.0%; Pred. No. 3.7e-36;  
 Matches 164; Conservative 91; Mismatches 253; Indels 99; Gaps 17;  
 QY 5 TLRTSAVLLTPAFAQVTP--ITDEL--ANPAGEWINGNOENYRHSPLTQI 59  
 DB 15 TLVAGTICALISGYATMASADGQATGEAIIHADHPGNMTYGRYSQDRYSPDQI 74  
 QY 60 TADNVGQLQVW-----AGMAGAVQVTPMTHDGMVTLANGDVICQLAQDGLDIME 113  
 DB 75 NRSNVGKLKAWLIDLTNRGEG--TPLVIDGMVATTNWSMKAVDAATGKLMS 129  
 QY 114 HRRLPA-VATLNAQGRKRGVALYGTSLYSSMDNHLALDMETQGVFV----- 164  
 DB 130 YDRPVENINADKGGCDPVNRGAAYMNGKVFGRFDRRLALDAKTLVWSVNTIPBEA 189  
 QY 165 ---ERGGEDGLTNTTGPVANG-VYVAGSTQVSPYFCFISGHSAAGEELMRNHF 220  
 DB 190 LGRKRSYTVVG-----APRIAGRVYIIGNGSEFGARG-FVSAFPAELGKXVWRFFTP 242  
 QY 221 QPGEEDG-----ETWGNDFEARMTG-----VMQGITVDPVNTLVFVG 258  
 DB 243 NPKKEPRAADSVIANKAYQVWS-----FTGAMTRQGGCGGVMSIYDPAADLVYIG 295  
 QY 259 STGVGPASETQGTGGTLYGNTREPAVDPDTEBIYMRHQTLEPRDNDQECTFEMNVANV 318  
 DB 296 VGNQSPNNYKRSRSGKDNLFISIVALKPDETEYVWHFQETPMQMDFTSDQINTLIDL 355  
 QY 319 DVQPSAMEGLRAINPAAATGERRVLTGAPCKTGTMWSPFAASGEELMRADTYNTMAS 378  
 DB 356 PI-----NGETRHVIVHARKNGFFIIDANTGEISGKNYYIVNMAAG 398  
 QY 379 ID-ETGLVTNEDAVLKEIDVEYDVCPFLGGRDWSAALNPDTGIYIPLNNACY--- 433

DB 399 LDPKTRPTVNDLTYLTGKWKVIGIPDGLGHNFAAMAFSPKGLVYITPAQGVPEFYTN 458  
 QY 434 -----DIYAVDOEFALDVYNTSATAKLAPGFENMRIDAIDISTRTLSAERPA 484  
 DB 459 QVGFTPHPDSDMNLGLDMKNKVIDP-SPEAKQAFVKDKMWIVAWDPQKQAEAMRVVDHK 517  
 QY 485 ANSPVLTAGVVPFNGTDRYFPALSOFTGETLMQARLATAVNTGQALSYELGVOYIA- 543  
 DB 518 FNGGILATGSDLLFQGLANGEFHAYDATNGSDLFFHFAADSGIITAPVYTLANGKQYVAV 577  
 QY 544 -IGAGGL 549  
 DB 578 EVGWGCI 584  
 RESULT 14  
 AAR05235  
 ID AAR05235 standard; protein; 742 AA.  
 XX  
 AC AAR05235;  
 XX  
 DT 24-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 04-AUG-1990 (first entry)  
 XX  
 DE Amino acid sequence of alcohol dehydrogenase (ADH) encoded by bases 1-  
 DE 2229 of alcohol dehydrogenase (ADH) gene.  
 XX  
 KM Alcohol dehydrogenase (ADH) gene; Acetobacter pasteurianus IFO 3191;  
 KM Acetobacter aceti K1006 (FERM-7528); plasmid PA4121.  
 OS Acetobacter aceti; strain K1006 (FERM-7528).  
 PN JP02000452-A.  
 PD 05-JAN-1990.  
 XX  
 PF 29-MAR-1988; 88JP-00075069.  
 PR 30-OCT-1987; 87JP-00273190.  
 PA (QPP) OP CORP.  
 PA (QPP) KEMPIE JOZO KK.  
 DR WPI; 1990-047990/07.  
 DR N-PEDB; AAQ91811.  
 XX  
 PT Cloning DNA, plasmid and microbe contg. it - contains alcohol  
 PT dehydrogenase gene derived from acetobacter aceti K10006 strain.  
 PS  
 PS Disclosure; Fig 4; 8pp; Japanese.  
 XX  
 CC Also new are a recombinant plasmid contg. its encoding DNA, and a  
 CC microorganism transformed with the plasmid. A DNA fragment was obtd. from  
 CC A. aceti strain K1006 and was transferred to an ADH-defective strain. The  
 CC resulting plasmid, PA4121, contg. its gene was inserted into A.  
 CC pasteurianus strain IFO 3191 by the triparent method. Its gene is useful  
 CC for improving Acetobacter culturing. (Updated on 25-MAR-2003 to correct  
 CC PF field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-  
 CC MAR-2003 to correct PA field.) (Updated on 24-OCT-2003 to standardise OS  
 CC field)  
 CC  
 XX  
 SQ Sequence 742 AA;  
 Query Match 16.0%; Score 490.5; DB 2; Length 742;  
 Best Local Similarity 25.7%; Pred. No. 3.3e-31;  
 Matches 152; Conservative 102; Mismatches 256; Indels 81; Gaps 17;  
 QY 11 AAVLLTPAFAQVTPITDEL--ANPAGEWINGNOENYRHSPLTQITADNVGQLQ 68  
 DB 23 AAALPYAVPRAAGQNTGSAIITHADHPENMLSYGRYSQRYSPDQINSNVGDLK 82

QY 69 LV-----WARGMAGAVQVTPMTHDGVTLANPQDVITQALDAQOTGDLIWEHRQQLPA-V 121  
 DB 83 LLEGYTLDTNRGQGA-----TPLVVDGIMVATNTWMSMEALDAATGTLMOQDPKVPGIN 137  
 QY 122 ATLNAQDRKRGVALYGTSLYFSSMDNHLIALDMETQGVVEDRGSGEDGLSNITGPI 181  
 DB 138 ADKQCCCTVNRGAGYNNKGFVWGTFFDGRVLAAADAKTKKAWAVNTTPADASLQKQSYIV 197  
 QY 182 ----VANVIVAGS-TCQYSPYGCFTSGHDSATGEELENNHFIPOGEEGD----- 227  
 DB 198 DGAIVYAKGVLILGNGGAEFGARG-FVSAPDAETGKLKMFRTVYNNKNEPDDAASDNL 256  
 QY 228 ----EWMGDFEARMWT-----GVWQIIVDPVTNLYFGSGVGVPASETGCTGGTL 277  
 DB 257 MKRAYIKWGP--KQAVRQGGGGTWDSDLVDPVSDLIYAVANGSPMNKYSSEGIGSN 314  
 QY 278 YGNTREPAVPDGTGEIYVRHQTLPRDNWDOECTFEMWVANVDYQPSAEMEGLRINPMA 337  
 DB 315 LFLGSIYALKPTEGCEYVWHFOATFMDQDYTSVQIITLMPYK----- 358  
 QY 338 TGE-FRVLTPAPCTGTMSFPAASGEFLMARDNTNMTASIDE-TGLVTVEDAVLKE 395  
 DB 359 -GEMRHVYHAP-KNGFFVYLDAKTGFLSGKNVYQNMANGLDPLTGREMYPDGLYTL 416  
 QY 396 LDVEYDVCFTFLGGRDWSAALNPDGTIYFLPLNNAACDI-----MAVDQ 440  
 DB 417 NGKFWYGIPEGLAHNMMAVAYSKTHLYIIPAHQIPFGYKNQVGKPHADSMNVGLDM 476  
 QY 441 EFSALDYNTSATAKLAPFENNCRIDAIDISTGRTLSAERPAANYSPVLTAGAVEN 500  
 DB 477 TKQNLPTPEARFAYIK---DLHGWLAMPVYKMEITWKIDHKGPMNGGILATGDDLFO 533  
 QY 501 GGDRYFRALSOETGETLMOARLATVATGQAISEYLDGVQYIA--IGAGSL 549  
 DB 534 GLANGFHAIDATNGSDLYKFDAGSGIILAPMTIYVNGQYVAEVWGSI 584

RESULT 15  
 ADD24941  
 ID ADD24941 standard; protein; 803 AA.  
 XX  
 AC ADD24941;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Escherichia coli glucose dehydrogenase.  
 XX  
 KW Directed genetic engineering; galactomannanase; reduced activity;  
 KW enhanced activity; xanthan gum production; suspension stability;  
 KW emulsion stability; temperature resistance; pseudoplasticity; amylase;  
 KW cellulase; extracellular protease; intracellular protease;  
 KW glucose dehydrogenase; enzyme.  
 XX  
 OS Escherichia coli.  
 XX  
 PN US2003036176-A1.  
 XX  
 PD 20-FEB-2003.  
 XX  
 PF 10-AUG-2001; 2001US-00927827.  
 XX  
 PR 28-MAR-2001; 2001US-0279493P.  
 XX  
 PA (BOWE/) BOWER S G.  
 PA (RAMS/) RAMSEIER T M.  
 XX  
 PI Bower SG, Ramseier TM;  
 XX  
 DR WPI: 2003-625389/59.  
 DR N-PSDB; ADD24915.  
 XX  
 PT New transformed cell or organism having reduced or enhanced activity of  
 at least one protein, useful for producing xanthan gum, which are useful

PT for providing formulations and properties.  
 PS Claim 1; SEQ ID NO 69; 135pp; English.  
 XX  
 XX The present invention relates to polypeptide and polynucleotide sequences  
 CC from Xanthomonas campestris which may be used for activity reduction or  
 CC enhancement using directed genetic engineering. A transformed cell or  
 CC organism having reduced or enhanced activity of at least one such protein  
 CC e.g. galactomannanase can be generated by disrupting the gene encoding  
 CC the protein. The activity of the protein is reduced by the presence of an  
 CC antisense nucleic acid sequence. The nucleic acid sequence of the gene  
 CC encoding the protein is a recombinant sequence having at least one  
 CC mutation as compared to the wild-type gene encoding the protein. The  
 CC transgenic cell or microorganism are useful for producing xanthan gum,  
 CC which are useful for providing formulations and properties, such as long-  
 CC term suspension and emulsion stability in alkaline, acid, and salt  
 CC solutions, temperature resistance, and pseudoplasticity. The present  
 CC sequence represents an enzyme relating to the present invention.  
 XX

Sequence 803 AA;  
 Query Match 13.4%; Score 412; DB 7; Length 803;  
 Best Local Similarity 23.8%; Pred. No. 1.1e-24;  
 Matches 179; Conservative 67; Mismatches 239; Indels 266; Gaps 29;

18 ABAFAQVTPITDELAMPAG-----EWINVGRNGBENHSHPLTQITADNV 64  
 DB 129 AEGALTQVAGIAGAVNRPADQAPAPAMPVDDDMAYGRSQGCRISPLQINRDNV 188  
 QY 65 GOLQVWARGMEAGAV-----QVTPMTHDGVNLYANPQDVITQALDAQOTGDLIWEH--- 114  
 DB 189 SOLQQAWV--FHTGVPFRKMGAEETPLKVGDSLYLCGARQVIALDASTQKQRRYDPK 246  
 QY 115 --RQQLPAVILNAQDGRKRGVALY-----GSLYFSSMDN 148  
 DB 247 VKDEAIPFYTAAC-----RGVSYEVEPTAAPAANDAAAGAAPALCRFRVIEGTLDD 299  
 QY 149 HLIALDMET-----GQVPEVERSGEDGILSTNTGPIYANGVIVAGSTC----- 193  
 DB 300 RLIALDARTGPRCADFGTHGVDTIVMGERTPPGVISINSPALYRGVVVIGHQVLDQCK 359  
 QY 194 QYSPYGCFTSGHDSATGEBLN-----RNHFIPOGEEGBETGNDPEAAMTGVWQC- 245  
 DB 360 RYBPSG-VIOGFDAVTGLRWAMDTPENMGAPPG---QVTRGTNPMTTAADEQ 414  
 QY 246 --ITVDPVTNLY--FYGSTGVGPASETQGTGPGTLVGTNTRFAYVRPDTGEIYVRHQTL 301  
 DB 415 LGVYVLEPMGNSTADYSSS-----RTPPENRYATSL-VALDVTGKPVANFQTH 463  
 QY 302 RDNWDQECTFEMWVANVDYQPSAEMEGLRINPMAATGEREVLTGAPCKGTWMSFPAAS 361  
 DB 464 IDAMDYD-----LGSQSLIDPFDGVNPAVL-----LPSKQELLYLDRRT 506  
 QY 362 GEFLMARDNTNMTASIDETGLTVNEDAV-----LKE 395  
 DB 507 GE-----PLVGEVERAVPGGVEPQKRYKQPSLYHTLTKPDLTE 547  
 QY 396 LD-----VEYDVC-----PTFLGARDWSAALNPDGTIY 424  
 DB 548 RDMWGMPRIDQLVGRICPRKASYKGIYTPPEADSHSIEYPPGNGSGDVGSAVDPHRCVI 607  
 QY 425 FLPLAN-ACYDIM-----ANDQFSALDYNTSATAKLA 457  
 DB 608 VANYNDMENYINILVPRAKADKLGNAPRDQVGDAGAGAGADPGAGTPYAVANVAGWFLP 667  
 QY 458 PGFENN-----GRIDAIDISTGRTLSAERP-----AANYSP 489  
 DB 668 --FTTLCKQPPYGGIRALIDKSGKITL--DRPPGSAKNGNPFGRISGLPIEIGTPNNGG 723  
 QY 490 VLSTAGVVP--NGGDRYFRALSOETGETLMOARLATVATGQAISEYLDGVQYIYALGAG 548  
 DB 724 SVTVAGGLIFFAAADVDIIRAIIDLATGKEIWHAKLPAGQANPMVYAVAGREYLVIMAAAG 783

Qy 549 LTYGTQJNAPLAEIDSTSVGNAIYFALPQ 579  
Db 784 -----HHFMETPAGDALVAYALPQ 802

Search completed: November 18, 2004, 05:16:09  
Job time : 165 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: November 18, 2004, 05:13:08 ; Search time 40 Seconds  
(without alignments)  
959,953 Million cell updates/sec

Title: US-10-802-682-8  
Perfect score: 3069  
Sequence: 1 MNPTLLRSLAAVLLTAPL.....AAIDSTSVGNALYVAPLPQ 579

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCITUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3069	100.0	579	US-09-470-667-8	Sequence 8, Appl1
2	2611	85.1	579	US-09-470-667-5	Sequence 5, Appl1
3	2526.5	82.3	578	US-09-470-667-6	Sequence 6, Appl1
4	2494	81.3	578	US-09-470-667-7	Sequence 7, Appl1
5	585	19.1	724	US-09-296-284-4	Sequence 4, Appl1
6	573.5	18.7	720	US-09-296-284-25	Sequence 25, Appl1
7	552	18.0	685	US-09-252-991A-32033	Sequence 32033, A
8	551.5	18.0	738	US-07-985-458-3	Sequence 3, Appl1
9	380.5	12.4	803	US-09-543-681A-4312	Sequence 4312, Ap
10	375.5	12.2	818	US-09-328-352-5208	Sequence 5208, Ap
11	367	12.0	740	US-09-635-145A-2	Sequence 2, Appl1
12	359.5	11.7	739	US-09-136-251-2	Sequence 2, Appl1
13	359.5	11.7	739	US-09-634-496-2	Sequence 2, Appl1
14	341.5	11.1	826	US-09-328-352-7515	Sequence 7515, Ap
15	302	9.8	602	US-09-489-039A-12436	Sequence 12436, A
16	286	9.3	688	US-09-489-039A-9813	Sequence 9813, Ap
17	216.5	7.1	705	US-09-252-991A-25730	Sequence 25730, A
18	161.5	5.3	243	US-09-252-991A-25814	Sequence 25814, A
19	154.5	5.0	339	US-09-543-681A-7743	Sequence 7743, Ap
20	133.5	4.3	335	US-09-540-236-2537	Sequence 2537, Ap
21	131	4.3	621	US-09-311-626B-2	Sequence 2, Appl1
22	126	4.1	715	US-09-710-279-570	Sequence 570, App
23	125	4.1	401	US-09-489-039A-11546	Sequence 11546, A
24	123	4.0	846	US-09-134-001C-3067	Sequence 3067, Ap
25	122	4.0	622	US-09-311-626B-4	Sequence 4, Appl1
26	122	4.0	1721	US-08-700-651-5	Sequence 5, Appl1
27	121.5	4.0	445	US-09-252-991A-22912	Sequence 22912, A

28	120.5	3.9	2736	4	US-09-252-991A-30227	Sequence 30227, A
29	120	3.9	910	4	US-09-710-279-3108	Sequence 3108, Ap
30	119.5	3.9	1042	3	US-08-928-361B-11	Sequence 11, Ap
31	119.5	3.9	1042	4	US-08-928-361B-5	Sequence 5, Appl1
32	119.5	3.9	1837	3	US-08-928-361B-5	Sequence 5, Appl1
33	119.5	3.9	1837	4	US-08-928-361B-5	Sequence 5, Appl1
34	119	3.9	1043	3	US-08-928-361B-30	Sequence 30, Appl1
35	119	3.9	1721	3	US-08-928-361B-6	Sequence 6, Appl1
36	119	3.9	1721	4	US-08-928-361B-6	Sequence 6, Appl1
37	118.5	3.9	470	4	US-09-710-279-2496	Sequence 2496, Ap
38	114	3.7	462	4	US-09-252-991A-29038	Sequence 29038, A
39	112	3.6	3567	2	US-07-642-734C-4	Sequence 4, Appl1
40	112	3.6	3567	3	US-08-439-009A-4	Sequence 4, Appl1
41	111	3.6	1183	3	US-09-134-001C-3530	Sequence 3530, Ap
42	110.5	3.6	383	3	US-09-134-001C-3068	Sequence 3068, Ap
43	110.5	3.6	1099	4	US-09-881-654-4	Sequence 4, Appl1
44	110.5	3.6	1099	4	US-10-637-323-4	Sequence 4, Appl1
45	110	3.6	3892	4	US-09-328-352-5503	Sequence 5503, Ap

## ALIGNMENTS

```

RESULT 1
US-09-470-667-8
; Sequence 8, Application US/09470667
; Patent No. 6730503
; GENERAL INFORMATION:
; APPLICANT: Asakura, Akira
; APPLICANT: Hoshino, Tatsuo
; APPLICANT: Ojima, Setsuko
; APPLICANT: Shinjo, Masako
; APPLICANT: Tomiyama, No. 6730503bunt
; TITLE OF INVENTION: No. 6730503el Alcohol/Aldehyde Dehydrogenases
; FILE REFERENCE: C38435/109700CON
; CURRENT APPLICATION NUMBER: US/09/470, 667
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: US 08/934,506
; PRIOR FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent version 3.1
; SEQ ID NO: 8
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Gluconobacter oxydans
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(23)
; OTHER INFORMATION:
US-09-470-667-8
Query Match 100.0%; Score 3069; DB 4; Length 579;
Best Local Similarity 100.0%; Pred. No. 2.8e-279;
Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MNPTLLRSLAAVLLTAPLPAFAQVTPITDELLANPPAGEMINNGRNDENTRHSPLTQIT 60
1 MNPTLLRSLAAVLLTAPLPAFAQVTPITDELLANPPAGEMINNGRNDENTRHSPLTQIT 60
1 MNPTLLRSLAAVLLTAPLPAFAQVTPITDELLANPPAGEMINNGRNDENTRHSPLTQIT 60
ADNVGQLVWARGMEAGAVOVTPKIHGVMVYLANPGVIVCALDAQTSDDLWERRQLPA 120
61 ADNVGQLVWARGMEAGAVOVTPKIHGVMVYLANPGVIVCALDAQTSDDLWERRQLPA 120
ADNVGQLVWARGMEAGAVOVTPKIHGVMVYLANPGVIVCALDAQTSDDLWERRQLPA 120
VATLNAQCDRRRGVALYGTSLYFSSGMDNHLTALDMETGQVVDYERSGEDGLTSNTTGP 180
121 VATLNAQCDRRRGVALYGTSLYFSSGMDNHLTALDMETGQVVDYERSGEDGLTSNTTGP 180
VATLNAQCDRRRGVALYGTSLYFSSGMDNHLTALDMETGQVVDYERSGEDGLTSNTTGP 180
IVANGVIYAGSTCOVSPYGCFTSGHDSATGELRNHFIPOGEGDPTWGNDEFEARMT 240
181 IVANGVIYAGSTCOVSPYGCFTSGHDSATGELRNHFIPOGEGDPTWGNDEFEARMT 240
IVANGVIYAGSTCOVSPYGCFTSGHDSATGELRNHFIPOGEGDPTWGNDEFEARMT 240
GWMGQITDPTNIVFYGSTVGPASSTFGTGTGTLVGNTRPAVRDTEIYWRQTL 300
241 GWMGQITDPTNIVFYGSTVGPASSTFGTGTGTLVGNTRPAVRDTEIYWRQTL 300

```

Db 241 GVMGQITVDVPTNLVHYGSTGVGPASSETGREGTLYGNTNFAVRPDTGELVWEHQTL 300  
Qy 301 PRDNWDECTFEMWVAVNDVQPSAEMEGRLAIPNNAATGERRVLTGAPCKGTGMSFDA 360  
Db 301 PRDNWDECTFEMWVAVNDVQPSAEMEGRLAIPNNAATGERRVLTGAPCKGTGMSFDA 360  
Qy 361 SGEFLMARDNTYNTMIAISIDETGLVTNEDAVLKEIDVEYDVCPTFLGGRDMSAALND 420  
Db 361 SGEFLMARDNTYNTMIAISIDETGLVTNEDAVLKEIDVEYDVCPTFLGGRDMSAALND 420  
Qy 421 TGIYFLPLNNAACVDYMAVDQPSALDYNTSATAKLAPGEMNGRIDAIDISTGRTLMSA 480  
Db 421 TGIYFLPLNNAACVDYMAVDQPSALDYNTSATAKLAPGEMNGRIDAIDISTGRTLMSA 480  
Qy 481 ERPAANSPVLSTAGGVVFNCGTDRYFRALSOETGETLWQARLATVATGQAISYEIDGQ 540  
Db 481 ERPAANSPVLSTAGGVVFNCGTDRYFRALSOETGETLWQARLATVATGQAISYEIDGQ 540  
Qy 541 YVIAAGGLTYGTOLNAPLA-EAIDSTSVGNATVVFALPQ 579  
Db 541 YVIAAGGLTYGTOLNAPLA-EAIDSTSVGNATVVFALPQ 579

## RESULT 2

US-09-470-667-5  
; Sequence 5, Application US/09470667  
; Patent No. 6730503  
; GENERAL INFORMATION:  
; APPLICANT: Asakura, Akira  
; APPLICANT: Hoshino, Tatsuo  
; APPLICANT: Ojima, Setsuko  
; APPLICANT: Shinjoh, Masako  
; APPLICANT: Tomiyama, No. 6730503ibumi  
; TITLE OF INVENTION: No. 6730503el Alcohol/Aldehyde Dehydrogenases  
; FILE REFERENCE: C38435/109700CON  
; CURRENT APPLICATION NUMBER: US/09/470,667  
; CURRENT FILING DATE: 1999-12-22  
; PRIOR APPLICATION NUMBER: US 08/934,506  
; PRIOR FILING DATE: 1997-09-19  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 579  
; TYPE: PRT  
; ORGANISM: Gluconobacter oxydans  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)..(23)  
; OTHER INFORMATION:  
US-09-470-667-5

Query Match 85.1%; Score 2611; DB 4; Length 579;  
Best Local Similarity 82.1%; Pred. No. 2,9e-236;  
Matches 476; Conservative 51; Mismatches 51; Indels 2; Gaps 2;

Qy 1 MNPFTLRTSAVALLTAPAPFAQVPTIDELLANPPAGMIVNGNOENYRHSPLTQIT 60  
Db 1 MKPTSLIMASAGALALAAPAPFAQVPTIDELLANPPAGMIVSYGNOENYRHSPLTQIT 60  
Qy 61 ADNVGQQLVWARGMEAGAVQVTPMIDHGVWYLANPGDVLOALDAQTGLIWEHRQLPA 120  
Db 61 TENVGQQLVWARGMQGKQVTPPLIHGVMYLANPGDVLOALDAQTGLIWEHRQLPN 120  
Qy 121 VATLNAQGRKRKGVALVGTSLYFSSWMDNHLIALDMETGOVFPVVERSGSDGLTNSITGP 180  
Db 121 IATLNSFGEPTRGVALVGTINVFVSWMDNHLVALDTATGQVTPVDRCQGED-MVSNSSGP 179  
Qy 181 IVANGVIVAGSTCOYSPYGCFIGSDGATGEELMRNHFIPOGREGDETGNDFEARMT 240  
Db 180 IVANGVIVAGSTCOYSPYGCFIGSDGATGEELMRNHFIIRAGREGDETGNDFEARMT 239  
Qy 241 GVMGQITVDVPTNLVHYGSTGVGPASSETGREGTLYGNTNFAVRPDTGELVWEHQTL 300

Db 240 GVMGQITVDVPTNLVHYGSTGVGPASSETGREGTLYGNTNFAVRPDTGELVWEHQTL 239  
Qy 301 PRDNWDECTFEMWVAVNDVQPSAEMEGRLAIPNNAATGERRVLTGAPCKGTGMSFDA 360  
Db 301 PRDNWDECTFEMWVAVNDVQPSAEMEGRLAIPNNAATGERRVLTGAPCKGTGMSFDA 360  
Qy 361 SGEFLMARDNTYNTMIAISIDETGLVTNEDAVLKEIDVEYDVCPTFLGGRDMSAALND 420  
Db 361 SGEFLMARDNTYNTMIAISIDETGLVTNEDAVLKEIDVEYDVCPTFLGGRDMSAALND 419  
Qy 421 TGIYFLPLNNAACVDYMAVDQPSALDYNTSATAKLAPGEMNGRIDAIDISTGRTLMSA 480  
Db 421 TGIYFLPLNNAACVDYMAVDQPSALDYNTSATAKLAPGEMNGRIDAIDISTGRTLMSA 479  
Qy 481 ERPAANSPVLSTAGGVVFNCGTDRYFRALSOETGETLWQARLATVATGQAISYEIDGQ 540  
Db 481 ERPAANSPVLSTAGGVVFNCGTDRYFRALSOETGETLWQARLATVATGQAISYEIDGQ 539  
Qy 541 YVIAAGGLTYGTOLNAPLA-EAIDSTSVGNATVVFALPQ 579  
Db 541 YVIAAGGLTYGTOLNAPLA-EAIDSTSVGNATVVFALPQ 579

## RESULT 3

US-09-470-667-6  
; Sequence 6, Application US/09470667  
; Patent No. 6730503  
; GENERAL INFORMATION:  
; APPLICANT: Asakura, Akira  
; APPLICANT: Hoshino, Tatsuo  
; APPLICANT: Ojima, Setsuko  
; APPLICANT: Shinjoh, Masako  
; APPLICANT: Tomiyama, No. 6730503ibumi  
; TITLE OF INVENTION: No. 6730503el Alcohol/Aldehyde Dehydrogenases  
; FILE REFERENCE: C38435/109700CON  
; CURRENT APPLICATION NUMBER: US/09/470,667  
; CURRENT FILING DATE: 1999-12-22  
; PRIOR APPLICATION NUMBER: US 08/934,506  
; PRIOR FILING DATE: 1997-09-19  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 579  
; TYPE: PRT  
; ORGANISM: Gluconobacter oxydans  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)..(23)  
; OTHER INFORMATION:  
US-09-470-667-6

Query Match 82.3%; Score 2526.5; DB 4; Length 579;  
Best Local Similarity 79.8%; Pred. No. 2.5e-228;  
Matches 462; Conservative 56; Mismatches 60; Indels 1; Gaps 1;

Qy 1 MNPFTLRTSAVALLTAPAPFAQVPTIDELLANPPAGMIVNGNOENYRHSPLTQIT 60  
Db 1 MKPTSLIMASAGALALAAPAPFAQVPTIDELLANPPAGMIVSYGNOENYRHSPLTQIT 60  
Qy 61 ADNVGQQLVWARGMEAGAVQVTPMIDHGVWYLANPGDVLOALDAQTGLIWEHRQLPA 120  
Db 61 TENVGQQLVWARGMQGKQVTPPLIHGVMYLANPGDVLOALDAQTGLIWEHRQLPN 120  
Qy 121 VATLNAQGRKRKGVALVGTSLYFSSWMDNHLIALDMETGOVFPVVERSGSDGLTNSITGP 180  
Db 121 IATLNSFGEPTRGVALVGTINVFVSWMDNHLVALDTATGQVTPVDRCQGED-MVSNSSGP 179  
Qy 181 IVANGVIVAGSTCOYSPYGCFIGSDGATGEELMRNHFIPOGREGDETGNDFEARMT 240  
Db 180 IVANGVIVAGSTCOYSPYGCFIGSDGATGEELMRNHFIIRAGREGDETGNDFEARMT 239  
Qy 241 GVMGQITVDVPTNLVHYGSTGVGPASSETGREGTLYGNTNFAVRPDTGELVWEHQTL 300



Db 240 GWMGQIITDYPVGLVHYGSSAVGPASETORGTGGTWTGINTREFAVRPEIGETIWMHQTL 299  
 QY 301 PRDWDQECTFEEMVAVNDVQPSAEMEGRLAIPNNAATGERRVLTGAPCKTGTMSFDAA 360  
 Db 300 PRDWDQECTFEEMVAVNDVQPSAEMEGRLAIPNNAATGERRVLTGAPCKTGTMSFDAA 359  
 QY 361 SGEFLMARDNTYNTMIASIDETGLVTNEDAVLKELDVEYDVCPTFLGGRDMSAALNDP 420  
 Db 360 TGEFLMARDNTYNTMIASIDETGLVTNEDAVLKELDVEYDVCPTFLGGRDMSAALNDP 419  
 QY 421 TGIYFLPNNACVDIMAVDOEFSSALDVYNTSATKLAPGFEEMGRIDAIDISTGRTLMSA 480  
 Db 420 TGIYFLPNNACVDIMAVDOEFSSALDVYNTSATKLAPGFEELVGRIDAIDISTGRTLMSA 479  
 QY 481 ERPAANYSPVLTAGVAVFNGGTDREYFRALSOETGETLMQARLATVATGAISYELDGVQ 540  
 Db 480 ERPAANYSPVLTAGVAVFNGGTDREYFRALSOETGETLMQARLATVATGAISYELDGVQ 539  
 QY 541 YVIAAGGLTYGTQNLNAPLAELDSTSVGNAYVYFALPQ 579  
 Db 540 YVIAAGGLTYGTQNLNAPLAELDSTSVGNAYVYFALPQ 578

RESULT 4  
 US-09-470-667-7  
 ; Sequence 7, Application US/09470667  
 ; Patent No. 6730503  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Asakura, Akira  
 ; APPLICANT: Hoshino, Tatsuo  
 ; APPLICANT: Ojima, Setsuko  
 ; APPLICANT: Shinjoh, Masako  
 ; APPLICANT: Tomiyama, No. 6730503Jbun1  
 ; TITLE OF INVENTION: No. 6730503el Alcohol/Aldehyde Dehydrogenases  
 ; FILE REFERENCE: C38435/109700CON  
 ; CURRENT APPLICATION NUMBER: US/09/470,667  
 ; CURRENT FILING DATE: 1999-12-22  
 ; PRIOR APPLICATION NUMBER: US 08/934,506  
 ; PRIOR FILING DATE: 1997-09-19  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 7  
 ; LENGTH: 578  
 ; TYPE: PRT  
 ; ORGANISM: Gluconobacter oxydans  
 ; FEATURE:  
 ; NAME/KEY: SIGNAL  
 ; LOCATION: (1)..(23)  
 ; OTHER INFORMATION:  
 ; US-09-470-667-7

Query Match 81.3%; Score 2494; DB 4; Length 578;  
 Best Local Similarity 79.3%; Pred. No. 2,9e-225; Indels 2; Gaps 2;  
 Matches 459; Conservative 56; Mismatches 62;  
 QY 1 MNPFTLLRTSAVALLTAPAAFAQVTPITDELLANPAGEMINYGNOENYRHSPLTQIT 60  
 Db 1 MKFTLLQSSAALLVLGTIPALAQ-TAITDEMLANPAGEMINYGNOENYRHSPLTQIT 59  
 QY 61 ADVNGOQLVYARMEGAVQVTPMIDGVMYLANPBDVIOALDAQCGDILMEHRQLA 120  
 Db 60 ADVNGOQLVYARMEGAVQVTPMIDGVMYLANPBDVIOALDAQCGDILMEHRQLA 119  
 QY 121 VATLNAQGRKRGVALYGTSLYFSSWDMNHLIALDMETGOVVFVERSGEDGLTSNTGP 180  
 Db 120 VATLNAQGRKRGVALYGTSLYFSSWDMNHLIALDMETGOVVFVERSGEDGLTSNTGP 178  
 QY 181 IVANGVYAGSTCOYSPYGCISGHDATGELMRNHFIPQGBEGDGTGNPPEARWMT 240  
 Db 179 IVANGVYAGSTCOYSPYGCISGHDATGELMRNHFIPQGBEGDGTGNPPEARWMT 238  
 QY 241 GWMGQIITDYPVGLVHYGSSAVGPASETORGTGGTWTGINTREFAVRPEIGETIWMHQTL 300  
 Db 240 GWMGQIITDYPVGLVHYGSSAVGPASETORGTGGTWTGINTREFAVRPEIGETIWMHQTL 299

Db 239 GWMGQIITDYPVGLVHYGSSAVGPASETORGTGGTWTGINTREFAVRPEIGETIWMHQTL 298  
 QY 301 PRDWDQECTFEEMVAVNDVQPSAEMEGRLAIPNNAATGERRVLTGAPCKTGTMSFDAA 360  
 Db 299 PRDWDQECTFEEMVAVNDVQPSAEMEGRLAIPNNAATGERRVLTGAPCKTGTMSFDAA 358  
 QY 361 SGEFLMARDNTYNTMIASIDETGLVTNEDAVLKELDVEYDVCPTFLGGRDMSAALNDP 420  
 Db 360 TGEFLMARDNTYNTMIASIDETGLVTNEDAVLKELDVEYDVCPTFLGGRDMSAALNDP 418  
 QY 421 TGIYFLPNNACVDIMAVDOEFSSALDVYNTSATKLAPGFEEMGRIDAIDISTGRTLMSA 480  
 Db 420 TGIYFLPNNACVDIMAVDOEFSSALDVYNTSATKLAPGFEELVGRIDAIDISTGRTLMSA 478  
 QY 481 ERPAANYSPVLTAGVAVFNGGTDREYFRALSOETGETLMQARLATVATGAISYELDGVQ 540  
 Db 480 ERPAANYSPVLTAGVAVFNGGTDREYFRALSOETGETLMQARLATVATGAISYELDGVQ 538  
 QY 541 YVIAAGGLTYGTQNLNAPLAELDSTSVGNAYVYFALPQ 579  
 Db 540 YVIAAGGLTYGTQNLNAPLAELDSTSVGNAYVYFALPQ 577

RESULT 5  
 US-09-296-284-4  
 ; Sequence 4, Application US/09296284A  
 ; Patent No. 6204040  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Choi, Bul-Sung  
 ; APPLICANT: Rhee, Sang-Ki  
 ; APPLICANT: Lee, Eun-Hae  
 ; TITLE OF INVENTION: Gluconobacter Suboxydans Sorbitol Dehydrogenase, Genes  
 ; TITLE OF INVENTION: and Methods of Use Thereof  
 ; FILE REFERENCE: 1533.0870000  
 ; CURRENT APPLICATION NUMBER: US/09/296,284A  
 ; CURRENT FILING DATE: 1999-04-22  
 ; NUMBER OF SEQ ID NOS: 87  
 ; SOFTWARE: Patent in Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 754  
 ; TYPE: PRT  
 ; ORGANISM: Gluconobacter suboxydans  
 ; US-09-296-284-4

Query Match 19.1%; Score 585; DB 3; Length 754;  
 Best Local Similarity 28.2%; Pred. No. 9e-46; Indels 106; Gaps 21;  
 Matches 172; Conservative 94; Mismatches 238;  
 QY 6 LRTSAV-LLTAPAAFAQ-VTPITDELLANPAGEMINYGNOENYRHSPLTQITAD 62  
 Db 16 LIGCAALAFCATSPVALAEDTGTATN--ADQHPGDMMSYGRYSBQRYSPLDQITKD 72  
 QY 63 NVGOQLVW-----ARGMEGAVQVTPMIDGVMYLANPBDVIOALDAQCGDILMEHR 116  
 Db 73 NASNKLAMHYDLITNRQEG---TPLIVGVVYATTNMSKKMLAALGKLMSTDP 127  
 QY 117 QLPV-VATLNAQGRKRGVALYGTSLYFSSWDMNHLIALDMETGOVVPDV----- 164  
 Db 128 KVPGNIALRGCDDVTNRGAAMNGKVFYGTBGRILALDATGKLVSVYVTPKXAGLGH 187  
 QY 165 ERSGEDGLTSNTPIYANG-VIYAGSTCOYSPYGCISGHDATGELMRNHFIPQGB 223  
 Db 188 QRSYTVDG-----APRIAKGVIIINGGAEGAGAG-FVYAYDAETGMDRFFTPVPD 240  
 QY 224 EGGD-----ETWGNDFEARWMTG---VWGQITDYPVNLVYFGSTGVGPASE 267  
 Db 241 NKPDGAADVDLMSKATYPTWKGAMKQGGGCTVMDLIDPVDVLYLVGNGSPNNY 300  
 QY 268 TORGTGGTGTNTREFAVRPEIGETIWMHQTLPRDWDQECTFEEMVAVNDVQPSA 327  
 Db 301 KFRSSGKGNLFLGSIIVAINPDTGKYVWFQETPPMDQMDYTSVOQIWA--LDMPVNGEM- 357  
 QY 328 GLRAINPNNAATGERRVLTGAPCKTGTMSFDAASEFLMARDNTYNTMIASIDE-TGLVT 386

Db 358 -----RHVYHAP-XNGFFYIIDATGKFIISGKPYTYENMANGLDPVTRGN 403  
QY 387 VNEDAVLKEDVEDVCPFTLGRDWSALNPDTGIFPLNNACYDIAMVDOESALD 446  
Db 404 YNDALMTLNGKRWYIGPDLGGHNFAMAYSPOTKLVYIPAOQVFF---VDPQKGGK 460  
QY 447 VNTSATKLAPEGHNGRIDAIDISTGR-----TLMSARPAANY-- 488  
Db 461 AHHDSWNLGL---DNKKGILLDDNDPOHKADKQFLDKGMIVAMPPOKQQAFTVD 515  
QY 489 -----PVLTAGGVFENGTDTRYPFALSOETGETLMQARLATVATGQAISEYLDGVQY 541  
Db 516 HKGPMNGGLATAGSVLFOGLANGEFHAYDATGKOLFTFPASALIAAPVITYTANGKQY 575  
QY 542 IA--IGAGGL 549  
Db 576 VAVEVGWGI 585

## RESULT 6

US-09-296-284-25  
Sequence 25, Application US/09296284A  
Patent No. 6204040  
GENERAL INFORMATION:  
APPLICANT: Choi, Eun-Sung  
APPLICANT: Rhee, Sang-Ki  
APPLICANT: Lee, Eun-Hae  
TITLE OF INVENTION: Glucobacter Suboxydans Sorbitol Dehydrogenase, Genes  
FILE REFERENCE: 1533.0870000  
CURRENT APPLICATION NUMBER: US/09/296,284A  
CURRENT FILING DATE: 1999-04-22  
NUMBER OF SEQ ID NOS: 87  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 25  
LENGTH: 720  
TYPE: PRT  
ORGANISM: Glucobacter suboxydans  
US-09-296-284-25

Query Match 18.7%; Score 573.5; DB 3; Length 720;  
Best Local Similarity 27.9%; Pred. No. 1e-44;  
Matches 164; Conservative 90; Mismatches 230; Indels 103; Gaps 19;  
QY 26 TPITDELLANPAGEMINYNQENYRHSPLTQITADNVQOLQVW-----ARGWEAGA 79  
Db 5 TAITN---ADQHGDMMSYGRYTSBQRYSPLDQITDMNSNLKLAHMYLDLTRGQEG-- 59  
QY 80 VQYTPMTHDGVNTLANBGVITQALDAQOTGLIWEHRRLPA-VATLNAQGRKGVALYIG 138  
Db 60 ---TPLLVDGVYATTNWSKKKALDAATGKLMSYDPKVPGNADRGCCDVTNRGAAYWN 116  
QY 139 TSLYFSSWDNLIALDEMGQVVDV-----ERQSGEDGLTNTGPIVANG-V 186  
Db 117 GKTYFGFDRLLALDKTKGLVMSYITVPEKAOLOHQSYYTDG-----APRIKGVY 170  
QY 187 IVAGSTQYSPYCGFISGHSATGEBELRNHFIPOGEEGD-----ETWQND 233  
Db 171 IIGGGAEGFARG-FVTAAYDAETGKMDWRFPTVFNPNKPKDGAASDVLMKAYPTWKGK 229  
QY 234 FEARMTMG---VMGQIYDPVTNLVFGSTGVGPASETQKGTGCTLYGNTFRFAVRPT 290  
Db 230 GAMQCGGGGTWDSLTLYDPVTLVYLGNGSPMNYKFBGKNNLFLGSLVALNPPT 289  
QY 291 GEIYWRHQLPRDNWDECTFEMNVAVNDVQPSAEVGLRATNPNAATGERVLTGAPCK 350  
Db 290 GKTYWHEQETPMQOWDYTSVQOIMA-LDWPVNGEM-----RHVYHAP-X 332  
QY 351 TGTMSDASGEFLMARDNTYNTMTASIDE-TGLVTVNEDAVLKEDVYDVCPTFLGG 409  
Db 333 NGFFYIIDATGKFIISGKPYTYENMANGLDPVTRGNPNDALMTLNGKRWYIGIPEDLGG 392

QY 410 RDMSSALNPDTGIFPLNNACYDIAMVDOESALVNTSATKLAPEGHNGRIDAI 469  
Db 393 HNFAMAYSPOTKLVYIPAOQVFF---VDPQKGGFRAHDSWNLG-----DNKKGILL 444  
QY 470 DISTGR-----TLMSARPAANY-- 504  
Db 445 DDNDPOHKADKQFLDKGMIVAMPPOKQQAFTVDHKGPMNGGLATAGSVLFOGLAN 504  
QY 505 RYFRLASOETGETLMQARLATVATGQAISEYLDGVQYIA--IGAGGL 549  
Db 505 GEFFAYDATGKDLFTFPASALIAAPVITYTANGKQYVAVEVGWGI 551

## RESULT 7

US-09-252-991A-32033  
Sequence 32033, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 32033  
LENGTH: 685  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32033

Query Match 18.0%; Score 552; DB 4; Length 685;

Best Local Similarity 27.2%; Pred. No. 9.6e-43;  
Matches 180; Conservative 110; Mismatches 230; Indels 142; Gaps 28;

QY 3 PTLTLETS-----AVILTPAFAQVPTITDELLAN--PPAGEMINYNQENYRHS 55  
Db 71 PAGILRPSHLCLAVAVLGSAGALLAK--DVTMEDINDKTKTGVDLYQWGTGAQRWSP 128  
QY 56 LTQITADNVQOLQVWMA-----RQMEGAVOYTMHNDGVYTLANBGDVITQALDAQ 107  
Db 129 LKQVADNVFPLTPMWSYSPFDEKQROESQAI-----VBDGYTVYVYASYSRLFALDAKT 183  
QY 108 GDLIWEHRRLP-----PAVATLNAQGRKGVALYGTSLYFSSWDNLIALDEMGQVVF 162  
Db 184 GKRLMTYHRLPDIIRCCDVVN-----RQALYDGKVFEGTLDASVALNKTKGVVW 237  
QY 163 DVERGSGEDGLTNTGPIVANG---VTVAGST-CQYSPYCGFISGHSATGEBELRN 216  
Db 238 KKKFADHAGAYTMGAPTIYADGKTGKLVLIHSGSGDEFVGLRFA-RPDPGEEELMR 296  
QY 217 HFIPOGEEG---DETWQNDPEA-----RMWTG---VMGQIYDPVTN 253  
Db 287 PFV---EGHMERLNGKDISTVIGDVAPSPMPPDDKNSPTKQVSMHGGGAPQOSAFDETN 354  
QY 254 LVFGSTGVGPASETQKGTGPG-----TLYGNTREAVRPTGEIYWRHQLPRDNWQ 307  
Db 355 TIIVAGNPGPMNWTARTAKGNPHDYDSLX-TSGQGVDPBSGGEVCMFYQHTPNDAMDF 413  
QY 308 ECTREMVAVNDVQPSAEMEGRLAINPNAATGERVLTGAPCKTGTMTMSDASGE--- 363  
Db 414 SGNNELVY-----FDYKADGKIVKATNAADRNGFFVYVDRSGKLQNA 457  
QY 364 -----ELMADTNTYNTMTASID-ETGLVTVNED--AVLKEDVEY---DYCPTFLGR 410  
Db 458 FPFVDNTIWA-----SHIDLKTRPYERBEQGRPLRPQKQKAVESPPFLGK 508  
QY 411 DWSSALNPDTGIFPLNNACYDIAMVDOESALDVNTSATKLAPEG-----ENM 463

Db 509 NNNMAYSDQTLGYFVPAH-----WKEDYWEBSYTKGSAIYLMGFRIRKMDH 561  
Qy 464 GRIDAIDISTRTLSAERPAANYSPLYSTAGVVFNGSDRYPRAISOENGETLMQARL 593  
Db 562 GSLAMPVSKVYWEKHEHPLMAGVLATNGNLVFTGTGDKYKADAKSGKELMFKQT 621  
Qy 524 ATVATGAISYELDVOY--IAIGAGLT--YG---TQNAPIAEALDSTSVGNALYFA 576  
Db 622 GSGIVSPITWEGQDEQYLGVTGAGVPLMGDMADLTRPVAQ-----GGSFVWFK 674  
Qy 577 LP 578  
Db 675 LP 676

RESULT 8  
US-07-985-458-3  
Sequence 3, Application US/07985458  
Patent No. 5344777  
GENERAL INFORMATION:  
APPLICANT: Tamaki, Toshimi;  
APPLICANT: Takemura, Hiroshi;  
APPLICANT: Tayama, Kenji;  
APPLICANT: Fukaya, Masahiro;  
APPLICANT: Okumura, Hajime and  
APPLICANT: Kawamura, Yoshiya  
TITLE OF INVENTION: Structural Gene of Membrane-Bound  
TITLE OF INVENTION: Alcohol Dehydrogenase Complex, Plasmid  
TITLE OF INVENTION: Containing The Same And Transformed Acetic Acid  
TITLE OF INVENTION: Bacteria  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fishauf, Holtz, Goodman & Woodward, P.C.  
STREET: 600 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10016-2088  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 0.72 mb  
COMPUTER: IBM PC compatible (NEC PC-9801 ES)  
OPERATING SYSTEM: MS DOS  
SOFTWARE: ASCII Form  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/985,458  
FILING DATE: 19921203  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/658,221  
FILING DATE: 20-FEB-1991  
APPLICATION NUMBER: 73440/1990  
FILING DATE: 26-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Goodman, Herbert  
REGISTRATION NUMBER: 17081  
REFERENCE/DOCKET NUMBER: 910134/HG  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 972-1400  
TELEFAX: (212) 370-1622  
TELEX: 236268  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 738 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: MATURE PEPTIDE  
LOCATION: 36 to 738  
IDENTIFICATION METHOD: N-terminal sequences of the  
IDENTIFICATION METHOD: purified protein having a molecular weight of about  
IDENTIFICATION METHOD: 72,000  
ORIGINAL SOURCE:

ORGANISM: Acetobacter alioaceti genes  
STRAIN: MH-24  
PUBLICATION INFORMATION:  
AUTHORS: Tamaki, Toshimi;  
AUTHORS: Takemura, Hiroshi;  
AUTHORS: Fukaya, Masahiro;  
AUTHORS: Tayama, Kenji;  
AUTHORS: Okumura, Hajime;  
AUTHORS: Kawamura, Yoshiya;  
AUTHORS: Nishiyama, Makoto;  
AUTHORS: Horinouchi, Sueharu and  
AUTHORS: Beppu, Teruniko  
TITLE: Cloning and Sequencing of the Gene Cluster  
TITLE: Encoding Two Subunits of Membrane-Bound  
TITLE: Alcohol Dehydrogenase from Acetobacter  
TITLE: polyoxogenes  
JOURNAL: Biochimica et Biophysica Acta.  
VOLUME: 1088  
PAGES: 292-300  
DATE: 1991  
US-07-985-458-3

Query Match 18.0%; Score 551.5; DB 1; Length 738;  
Best Local Similarity 27.0%; Pred. No. 1,2e-42;  
Matches 164; Conservative 91; Mismatches 253; Indels 99; Gaps 17;

Qy 5 TLRTSAVLLITAPAAQVTP---TDELL--ANPAGWINYGRNOENYRHSPLTQI 59  
Db 15 TTAGTICALLISGATMASADGQAGTGEALIHADHPGMNMTYGRYSDDQYSPLDQI 74  
Qy 60 TADNVGOLQVW-----ARGWEAGAVQVTPMHGQWYLANPGVIALDAQDQDILME 113  
Db 75 NRSNVGNLKLANYLLDITNGQEG---TPVITGVMTATTNMSMKAVDATTGKLLMS 129  
Qy 114 HRRQLPA-VATLNAQDRKRGVALYGTSLYFSSMDNHLIADMETGVVFDV----- 164  
Db 130 YDPRVPGNTADRGCCDVTNRGAAYNKGYYFGTDFGRLLIADAKTKLVMSVNTIPRAE 189  
Qy 165 ---ERSGSDGLTSMITGPVANG--VYVAGSTCOQSPFGCSFGSDATGELMNHFTIP 220  
Db 190 LCKQRSYIVDG-----APRLKGRVITGNGSEFGARG-FVSAPDAETGKDMKFFTPV 242  
Qy 221 QGEEBGD-----ETWGNDFEARMTG-----WGQITVDPVTLNLFVEYG 258  
Db 243 NPKNEPDAAASDVLNKAYQTWS-----PTGATRQGGGTVWDSIVYDPAVLVYL 295  
Qy 259 STGVGPASETQGTGFTGTYLGTNTPFAVRPTGELVMEHQTLPRDNWQECTEEMVANY 318  
Db 296 VANGSPWNYKYSSEKGNLFLGSIVALKPEYGEVYMHFOETPMQOWDFTSDQJINTIDL 355  
Qy 319 DVQPSAEMGLAIPNNAATGERAVLTGAPCTGTWMSFDDAASGEFLMARDNTNMTIAS 378  
Db 356 PL-----NGETRAVIYHARANGFFIITDAGTSFGISKIVYVNMASG 398  
Qy 379 ID-ETGLVTNEDAVLKEIDVEYDVCPTFLGGRDMSAALPDTGIYPLNNACY---- 433  
Db 399 LDPKTRGPPIYNDALYTLTGKEMYGIPDLGGHNPAAMAFSPKTLIVIPAAQVFLYTN 458  
Qy 434 -----DIAVOEFSALDVNTSATKLAPFEMNGRLDADISTGTWMAEPA 484  
Db 459 QVGGFTPHPDSWNLGLDMNKVGIPD-SPRAKAFYKDKMTIVADPOKQAEAMRVDKG 517  
Qy 485 ANYSPVLTAGGVVFNNGTDRYFRALSOETGETLMQARLATVATGAISYELDVOYIA- 543  
Db 518 PMNNGILATGDDLFLQGLANGEFHAYDATNGSDLHPFAADSGIAPVTVLANGQYAV 577  
Qy 544 -IGAGTL 549  
Db 578 EVGWGGI 584

RESULT 9  
US-09-543-681A-4312

; Sequence 4312, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543.681A  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 4312  
; LENGTH: 803  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
; US-09-543-681A-4312

Query Match 12.4%; Score 380.5; DB 4; Length 803;  
Best Local Similarity 24.5%; Pred. No. 1.6e-26;  
Matches 182; Conservative 87; Mismatches 240; Indels 233; Gaps 40;

QY 5 TLRTSAVLL--LTPAA--FAQVPTIDELLANPPAG--EWINGENQENYR 53  
DB 124 TLATTLIIIGSLTDPHAINQFTQCP--TNSI--SPIGVNDMPAYGRTQAGIRY 179  
QY 54 SPLTQITADNVGLOLVMA-----RGEAGAV--QVTPEHIDGVYLANPGDVIQALD 104  
DB 180 SPLQIINSNDVDKVEKMTHTGDERKANDAVEITNEVTIKIEDNLFCTPHQYISLD 239  
QY 105 AQTGDLMEHRRQIPAVATLNAQGRKRGVALYGTG-----L 141  
DB 240 PATGEKMRPDSKIQ--YNSKSFQHMTCRGVAYVYNNNSAEFAKSLDVRQPSIDCEPEKY 297  
QY 142 YFSSMDHLALDMETQGVVFDVERSGEDGLTS-----NTTGP--IYANVIYA 189  
DB 298 FLVPNDRLIAVNAQTGNACRDRGQ--NGEINILASMPYALGGYNTSPITVIGNTIIG 356  
QY 190 GSTQC---YSPYCFISGHSATGEEIMNHPFPGEGE-----GDETGNDFEARWM 239  
DB 357 GSVTDNLSTQEPSC-VIRGYDVNSGKLIW---VFDTGADPNAMGKET--QF-VHNS 407  
QY 240 TGWVGQITDPVNTLVFYSGTGVGPASGTQGTG--GGTLXGNTNR--AVRPDGG 291  
DB 408 PNANAPLAYDAQIDVV-YIPTGV-----GTPDIWGDGRHPLKERYANSVLALGSGTG 458  
QY 292 EIVVRHQTLPKDMWDOCTFEMVAVNDVQPSA-----EMEGIRAINPMAATGERRVL- 344  
DB 459 ELIMHFGTTHHDLMDM-----VPSQPSLVDIKTDKNETIPALYITKTGNVFLD 509  
QY 345 -----TGAPCKTGTMMS-----FDAS-----GEFLMA----- 367  
DB 510 RRDGTPIVPYTERKVPFVSKYKQPTQGEHNSPTQFSAIMLAPRNKLGSDMNGMTADQ 569  
QY 368 -----RDNYTMASIDETGLVTVNEDAVLKELDVEYDVCPFTLGGRDWSAALNDP 420  
DB 570 LACIAHQAUNBEIYTPPSBNGTL-----VFPGLGIFPFWGIGVND 613  
QY 421 TGI-----YFLPLNNACY-----DIMAADQ--EFSALDVYNTSATKLAPEGEMNR- 465  
DB 614 QGVAVNMFITLPFISKYLLKDPQDNQALDAGHGQGLQPMGTGTPYSVDIHFPISPLGLPC 673  
QY 466 -----IDMIDISTGTLMS-----AERP-----AANYSPVLSAGGVN 500  
DB 674 KQPMWGVAVGVDLTKNELVWQQRFTGRDLSLNLVLEPLKIGVPGIGGAIISAGNVNRY 733  
QY 501 GGT-DRYFRALSOFTGETLMOARLATVATGQA--ISYELDGVQYIAGAGL--TYGTQIN 556  
DB 734 AASGDNVIIRAFNVNTNGMLWEARLP--AGGQATPMTYISGKQYVIVMAGHGSPGKX- 790  
QY 557 APLAALDSTSVGNAIYVFPALP 578  
DB 791 -----GSLVAYGLP 800

RESULT 10  
US-09-328-352-5208  
; Sequence 5208, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328.352  
; PRIOR FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5208  
; LENGTH: 818  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
; US-09-328-352-5208

Query Match 12.2%; Score 375.5; DB 4; Length 818;  
Best Local Similarity 24.1%; Pred. No. 4.9e-26;  
Matches 177; Conservative 87; Mismatches 253; Indels 217; Gaps 38;

QY 6 LLRTSAVLLTAPAAFAQVPTIDELLANPP-----AGEWINGENQENYR 55  
DB 138 LSSTLAIIVLMVYSIFNDPQIINGEIKTPQETAGAVPGVABSDMPAYGRTQAGERYSP 197  
QY 56 LTOITADNVGLOLVMA-----RGEAGAV--QVTPEHIDGVYLANPGDVIQALDQ 106  
DB 198 LKQINDQVNDKQVAMFTRTGDFKTDNDSETTQVPIKGNMFTCTAHQQLAIDPA 257  
QY 107 TGDILMEHRRQIPAVATLNAQGRKRGVALY-----GTSL-----YFS 144  
DB 258 TGEKMRPDSKIQ--YNSKSFQHMTCRGVAYVYNNNSAEFAKSLDVRQPSIDCEPEKY 315  
QY 145 SWDNHLALDMETQGVVFDVERSGEDGLTS-----NTTGPVANG--YIVAGS- 191  
DB 316 VNDGLVAVNADTGACTDFGQ--NGEVNLDGFMPYALPGYNTSPITVIGNTIIGSV 374  
QY 192 TCQYS---PYGCFISGHSATGEEIMNHPFPGEGE-----GDETGNDFEARWM 240  
DB 375 TDNYSNKEPSC-VIRGYDVNTGKLIWFDGADPN--AMPGB-----GTFP-VHNSP 423  
QY 241 GVGQITDPVNTLVFYSGTGVGPASGTQGTG--GGTLXGNTNR--AVRPDGG 290  
DB 424 NAMAPLAYDAQIDVV-YIPTGV-----GTP--DIWGDRTLEKERYANSVLALGSGTG 458  
QY 291 EIVVRHQTLPKDMWDOCTFEMVAVNDVQPSAEMEGIRAINPMAATGERRVL--NAATGER- 342  
DB 473 GKLIMHFGTTHHDLMDMVPQPSLADIKKSGKTVAIYVLTGTGNAFVLDRNGQPIV 532  
QY 343 -----VLTGAPCKTGTMMS-----FDASGEFLMA-----R 368  
DB 533 FYTERKVPQTVKRGQPTQGEHNSKTQFSDNLNAPQKLDKDMGATMLDQLMCRVFSK 592  
QY 369 DNYTMASIDETGLVTVNEDAVLKELDVEYDVCPFTLGGRDWSAALNDP----- 420  
DB 593 RLNDGCIPTPSBNGTL-----VFPGLGIFPFWGIGVND 613  
QY 421 TGIYF---LPLN-NACYDIADVQFSAADV---YNTSATYADP-GF--ENMGRI 466  
DB 617 IGLPVSRLIPADPRAQAKAGTQGVQPMYGVVPGVEISATLSLGLPCXOPAWGVY 696  
QY 467 DAIDISTGRTLSA-----ERPA-----NYSPIVLSAGGVNNGGT-DRYF 507  
DB 697 AGVDLKTHEVWVKRIGITRDSLPNLQPLPVPVKIGVGLGSGISTAGNVNMFVATQNYI 756  
QY 508 PALSOFTGETLMOARLATVATGQA--ISYELDGVQYIAGAGL--TYGTQIN 556  
DB 757 RAFNVNBEKLMWARLP--AGGQATPMTYISGKQYVIVMAGHGSPGKX----- 805  
QY 565 STSVGNAIYVFPALP 578

```
Db      806 -----GDYLVAYALP 815

RESULT 11
US-09-635-145A-2
/ Sequence 2, Application US/09635145A
/ Patent No. 6653115
/ GENERAL INFORMATION:
/ APPLICANT: HOSHINO, Tatsuo
/ APPLICANT: MIYAZAKI, Taro
/ APPLICANT: OJIMA, Setsuko
/ APPLICANT: SHINJOH, Masako
/ APPLICANT: TOMIYAMA, No. 6653115bun1
/ TITLE OF INVENTION: D-SORBITOL DEHYDROGENASE GENE
/ FILE REFERENCE: C38435/118291
/ CURRENT APPLICATION NUMBER: US/09/635,145A
/ PRIOR FILING DATE: 2000-08-08
/ PRIOR APPLICATION NUMBER: EP 97114432.4
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 2
/ LENGTH: 740
/ TYPE: PRT
/ ORGANISM: Glucobacter suboxydans
/ FEATURE:
/ NAME/KEY: SIGNAL
/ LOCATION: (1)...(24)
US-09-635-145A-2

Query Match      12.0%; Score 367; DB 4; Length 740;
Best Local Similarity 23.7%; Pred. No. 2.6e-25;
Matches 171; Conservative 83; Mismatches 230; Indels 238; Gaps 35;

Qy      23 AOVPTITDELLANPAGEMINYGROENYRHSPLTQITADNVGQLOL--VWARG----- 74
Db      86 SQVPAMAPQGSANPARGDWVAYGRDHDQTRYSPLSEITPENASKUKVAFVHTGSPRRG 145
Qy      75 -MEAGAVQVTPM-IHDGVMTLANPQDVYQALDAQDGLIWEHR-----ROLPAVATLNAQ 127
Db      146 QVNKMAETTPIKVGDGLYTCSANMDIIR-LDPATGKQIRRNVDVKXHSIPYTAAC--- 201
Qy      128 GDRKRGVALYGS-----LYSSMDNHLIADMETGQV-----FDVERGS 168
Db      202 ---KGVYFTSSVVPBGQPCNRLIEGTLDMRLIADVAFGDCPCNFGHGQVNLMOGL 257
Qy      169 GED--GLTSNTTGPVANGVIAGSTC-----QSPYGCFTSGHDSATGEELW---RNH 217
Db      258 GESVPGFVSMTAPRPVINGVVVNHVELDQGRNAPSG-VIRGYDAESGKFWAMDVNNS 316
Qy      218 FIPOGGEDETWGNDPEARMGTGWCQITVDPTNLVFGSTGVGPASSETQRTPGGTL 277
Db      317 GRSQPA-----YRVTVITAVERRIP-----GLPDRRQRCG----- 345
Qy      278 YGTNTR-----FAVRPD-----TGEIWRHQTLPRDNWDOECTFE 312
Db      346 -GSRLRDRNSADYYSALRSDAENKYSAAVAIDVKTGSPRWVFGTAHKDQVMDYDISQ 404
Qy      313 ---MVAVANDVQ-----PSAENEGGLAINPNAATGERRVL- 345
Db      405 ATLMDMPGPDQVPAALIMPTKRGQTFVLDRRTGKPIIPVEERPAPSPGVIIPGDRSPFQ 464
Qy      346 ---GAPC-----KGTGWSFDAASGEF--LWARDNYNNMIA--SIBETGLVTYNEBA 391
Db      465 PWSGPMALAVPDIKFTDMGMSPIDQLFCRIKFRANVGETTPPSVDKPM----- 516
Qy      392 VLKELDVEYDVCPFTFLGDRMSSAALNPDTGIYFLPLN-NACVD-----INAV 438
Db      517 ---IEY---PQYNGSDWGSMSYDPQSGILIAMNNTPTWVDQVTRKKADSLGLMEI 567
Qy      439 D-----QFSLADVYNNSATAKLAFGEN-----MGRIDALDISG-RT 476
```

```
Db      568 DDNPFKGGGAGENGAMVD--GTPYGIIVTPWMDQYTGKMCNRPYGMITALDMKHGXV 625
Qy      477 LWS-----AERP-----AANYSPVLSTAGGVVNG--GTDYRFPALSGEIGE 516
Db      626 LWQHPGLTARANGPGLPTGLPWEIGTPNNGSGSVTGGGLPIGATDQJAIIDEHQK 685
Qy      517 TLWQARLATVATGATISVELDGVYIATGAGGLVTGTOLNAPLAABDSTSVGNALVYA 576
Db      686 VVMSAVLPBGQGANPMTYHANGHQYVALMAGS-----HHFMPTVSDQLVYIA 733
Qy      577 LP 578
Db      734 LP 735

RESULT 12
US-09-136-251-2
/ Sequence 2, Application US/09136251A
/ Patent No. 6127156
/ GENERAL INFORMATION:
/ APPLICANT: HOSHINO, Tatsuo
/ APPLICANT: MIYAZAKI, Taro
/ APPLICANT: OJIMA, Setsuko
/ APPLICANT: SHINJOH, Masako
/ APPLICANT: TOMIYAMA, No. 6127156bun1
/ TITLE OF INVENTION: D-SORBITOL DEHYDROGENASE GENE
/ FILE REFERENCE: D-Sorbitol Dehydrogenase Gene
/ CURRENT APPLICATION NUMBER: US/09/136,251A
/ PRIOR FILING DATE: 1998-08-19
/ EARLIER APPLICATION NUMBER: EP 97114432.4
/ EARLIER FILING DATE: 1997-08-21
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 739
/ TYPE: PRT
/ ORGANISM: Glucobacter suboxydans
/ FEATURE:
/ NAME/KEY: SIGNAL
/ LOCATION: (1)...(24)
US-09-136-251-2

Query Match      11.7%; Score 359.5; DB 3; Length 739;
Best Local Similarity 23.7%; Pred. No. 1.3e-24;
Matches 171; Conservative 83; Mismatches 229; Indels 239; Gaps 36;

Qy      23 AOVPTITDELLANPAGEMINYGROENYRHSPLTQITADNVGQLOL--VWARG----- 74
Db      86 SQVPAMAPQGSANPARGDWVAYGRDHDQTRYSPLSEITPENASKUKVAFVHTGSPRRG 145
Qy      75 -MEAGAVQVTPM-IHDGVMTLANPQDVYQALDAQDGLIWEHR-----ROLPAVATLNAQ 127
Db      146 QVNKMAETTPIKVGDGLYTCSANMDIIR-LDPATGKQIRRNVDVKXHSIPYTAAC--- 201
Qy      128 GDRKRGVALYGS-----LYSSMDNHLIADMETGQV-----FDVERGS 168
Db      202 ---KGVYFTSSVVPBGQPCNRLIEGTLDMRLIADVAFGDCPCNFGHGQVNLMOGL 257
Qy      169 GED--GLTSNTTGPVANGVIAGSTC-----QSPYGCFTSGHDSATGEELW---RNH 217
Db      258 GESVPGFVSMTAPRPVINGVVVNHVELDQGRNAPSG-VIRGYDAESGKFWAMDVNNS 316
Qy      218 FIPOGGEDETWGNDPEARMGTGWCQITVDPTNLVFGSTGVGPASSETQRTPGGTL 277
Db      317 GRSQPA-----YRVTVITAVERRIP-----GLPDRRQRCG----- 345
Qy      278 YGTNTR-----FAVRPD-----TGEIWRHQTLPRDNWDOECTFE 312
Db      346 -GSRLRDRNSADYYSALRSDAENKYSAAVAIDVKTGSPRWVFGTAHKDQVMDYDISQ 404
Qy      313 ---MVAVANDVQ-----PSAENEGGLAINPNAATGERRVL- 345
Db      405 ATLMDMPGPDQVPAALIMPTKRGQTFVLDRRTGKPIIPVEERPAPSPGVIIPGDRSPFQ 464
```

```
QY 346 -----GAPC-----KGTGMSFPAASGEF--LMARDNTYNTMIA--SIDETGLVTVNEDA 391
DB 465 PMSVGMFALRVPLDKETDMGMSPIDOLFRIKFRANVGEFTPPSVDXPM-----IMAV 516
QY 392 VLKELDEYDVCPFLGGRDMSAALNPDGTGYELPLN--NACVD-----IMAV 438
DB 517 -----IEY-----PGYNGSDMSMSYDPQSGILIANMNTIPMDQVTRKKADSLGLMFI 567
QY 439 D-----OEFALDVTNTSATKALAPFEN-----MGRIDALDISTG-RT 476
DB 568 DDNFKFGGGAGENGAMD--GTPYGIIVTPFMDQYTMKCNRPYGMITAIIDMKHQXV 625
QY 477 LMS-----AERP-----AANYSPVLSTAGGVFENG-GTDRYFRALSOETGE 516
DB 626 LMOHPLGTARANGPWGLPTGLPWEIGTPNNGSSVVTGGILIFGATIDNOIRALDEHTGX 685
QY 517 TLMQARLATVATGQATSYELDGVQYIAGAGLTYGTQNLAPLAELIDSTVGNALVFA 576
DB 686 VVMSAVLPFGGGQANPMTYEANGHQYVAL-AGG-----HHFMTPVSDQLVVA 732
QY 577 LP 578
DB 733 LP 734
```

## RESULT 13

```
US-09-634-496-2
/ Sequence 2, Application US/09634496
/ Patent No. 644449
/ GENERAL INFORMATION:
/ APPLICANT: HOSHINO, Tatsuo
/ APPLICANT: MIYAZAKI, Taro
/ APPLICANT: OJIMA, Setsuko
/ APPLICANT: SHINJOH, Masako
/ APPLICANT: TOMIYAMA, No. 644449:ibumi
/ TITLE OF INVENTION: D-SORBITOL DEHYDROGENASE GENE
/ FILE REFERENCE: D-Sorbitol Dehydrogenase Gene
/ CURRENT APPLICATION NUMBER: US/09/634,496
/ EARLIER FILING DATE: 2000-08-08
/ EARLIER APPLICATION NUMBER: US/09/136,251
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 739
/ TYPE: PRT
/ ORGANISM: Gluconobacter suboxydans
/ FEATURE:
/ NAME/KEY: SIGNAL
/ LOCATION: (1)..(24)
US-09-634-496-2
```

```
Query Match 11.7%; Score 359.5; DB 4; Length 739;
Best Local Similarity 23.7%; Pred. No. 1.3e-24;
Matches 172; Conservative 83; Mismatches 229; Indels 239; Gaps 36;

QY 23 AQTPTIDELLANPBAEMINVGNGNENYHSPLOTANVQQLQ--VMAG-----74
DB 86 SQVPAMAPQOSANPARQDWAAYGRDDHQTRYSPSETPENASKLVAFYHTGTSYPRQG 145
QY 75 -MEAGVQVTPM-IHDSVMTLANPQVIALDAQTGDLWEHR-----ROLPAVALTLNQ 127
DB 146 QVNWMALETPPIKVGDDLYTCSANNDIK-LDPATGQIWRNVVDVYKISHIPTAAC--201
QY 128 GDRKGVALYGT-----LYSSWDNHLIALDMETGOV-----FVDERGS 168
DB 202 ---KGTTFYFTSSVPEGQPCNRLIEGLDMRLIADVAETGDFCFNGHGGQVNLMOGL 257
QY 169 GED--GTSNTTGTIVANGVIVAGSTC---QYSPYGCFTISGHDSATGEELW---RNH 217
DB 258 GESVPGSVKTAAPPVINGVIVVNHVLDQGRMAWPG-VIRGVDAESGKFWAMVNVNS 316
```

```
QY 218 FIPQGEEDETWGNDEFAARMGTGVMQOITYDPVTLVFFGSGTGVPASFTORGTPGTL 277
DB 317 GRSQPA-----YRVVITAVERRIP-----GLPDRRGG-----345
QY 278 YGINTR-----FAVRPD-----TGEIWMRHQTLPRDMNDQECTFE 312
DB 346 -GSRLEPRDRNSAADYSALRSDAENKVSAAVVAIDVTGSPRVVFOIAHXDVVDYDGSQ 404
QY 313 ---MVAIVADVQ-----PEAMEGLRAINPAATGERVLT- 345
DB 405 ATLMDMPGPDQGYPALIMPITKGGQTFVDRRTGKPLPVEERPAAPSPGVPIDPSPTQ 464
QY 346 ---GAPC-----KGTGMSFPAASGEF--LMARDNTYNTMIA--SIDETGLVTVNEDA 391
DB 465 PMSVGMFALRVPLDKETDMGMSPIDOLFRIKFRANVGEFTPPSVDXPM-----IMAV 516
QY 392 VLKELDEYDVCPFLGGRDMSAALNPDGTGYELPLN--NACVD-----IMAV 438
DB 517 -----IEY-----PGYNGSDMSMSYDPQSGILIANMNTIPMDQVTRKKADSLGLMFI 567
QY 439 D-----OEFALDVTNTSATKALAPFEN-----MGRIDALDISTG-RT 476
DB 568 DDNFKFGGGAGENGAMD--GTPYGIIVTPFMDQYTMKCNRPYGMITAIIDMKHQXV 625
QY 477 LMS-----AERP-----AANYSPVLSTAGGVFENG-GTDRYFRALSOETGE 516
DB 626 LMOHPLGTARANGPWGLPTGLPWEIGTPNNGSSVVTGGILIFGATIDNOIRALDEHTGX 685
QY 517 TLMQARLATVATGQATSYELDGVQYIAGAGLTYGTQNLAPLAELIDSTVGNALVFA 576
DB 686 VVMSAVLPFGGGQANPMTYEANGHQYVAL-AGG-----HHFMTPVSDQLVVA 732
QY 577 LP 578
DB 733 LP 734
```

## RESULT 14

```
US-09-328-352-7515
/ Sequence 7515, Application US/09328352
/ Patent No. 6562958
/ GENERAL INFORMATION:
/ APPLICANT: Gary L. Breton et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
/ FILE REFERENCE: GTC99-03PA
/ CURRENT APPLICATION NUMBER: US/09/328,352
/ EARLIER FILING DATE: 1999-06-04
/ NUMBER OF SEQ ID NOS: 8252
/ SEQ ID NO 7515
/ LENGTH: 826
/ TYPE: PRT
/ ORGANISM: Acinetobacter baumannii
US-09-328-352-7515
```

```
Query Match 11.1%; Score 341.5; DB 4; Length 826;
Best Local Similarity 23.4%; Pred. No. 7.8e-23;
Matches 157; Conservative 83; Mismatches 215; Indels 217; Gaps 28;

QY 41 WINGRNENRHSPLQITADNVGOLVW-----AGMEAGAV-QYTPMIDGVMY 92
DB 190 WENGENDGGSRFYALDDINRDNVHKLEAKRFRFTGFTTGSNGAGADQMTPLDVGKVL 249
QY 93 LANPQDVIALDAQTGDLWEHRROLPAVALTLNAQD--RKRGVALYGT-----140
DB 250 LCPHNNIFALDADSGKQIWK-----AEVNSKADAMERCAGVAFDSTKPLVQPTLA 301
QY 141 -----LYSSWDNHLIALDMETGOVVF-----VERSGEGGLTSN- 176
DB 302 GATPVTVAANTACPRRYTNTPGRLIIVANVADGQSCADFGVNGVYDLLEGLOGGSKAP 361
QY 177 ---TTGPVIVANGVIVAGS---TCQYSPYGCFTISGHDSATGEELW---RN--HFIPO 221
```

```

Db      362 REEVTAETIAGTTIVGSRIDNVAADMGGVIRGVYITGLKLMAPDRPNDDPNVYK 421
Qy      222 PGESEDETGNDEAFARMWTGVMGQITYPVTLNLYF--YGSTGVPAASETOGFGTLYG 279
Db      422 PGEIYKSSANS-----MAMSYDPQNTYVFLMSSSSVDINKGN--NPLDKYN 470
Qy      280 TMTFAVRPDTGELVMEHQTLPRDNMQOECTFEMKVANVDVQSAEVEGLPAINPNAATG 339
Db      471 TSV-LADATGKKEKMYQIVHNDLMD-----FDLWQPSLVDPLK----DGIT 515
Qy      340 ERRLTGAPCKTGTMTGSEFDAASGEFLWABDTNNTMI-----AS 378
Db      516 KPAVVIQT--KSGGFPLVDRVTGKPL-----TKVLEQPKAANTGEOYSLTOPSVE 566
Qy      379 IDETGLVTNEDAVLKELDVEYDYC-----PTFLGGRDMS 414
Db      567 MPOIGNOTLTRESDMGATPFDQLMCRINFKSMRYEGFLTAPGTDVSLSPGSLGMMWS 626
Qy      415 AALNP-----DGIYFLPANNACYDIMAY 438
Db      627 IAPPTTRRYFVNDMLGLWVQLIKQTPEDIKIQASGGEKVTGMAVPMKGPYKVNK- 685
Qy      439 DOERSALDVNTSATKLAPGFENNGRIDAIDISTGRTLSAERPAAN-----Y 487
Db      686 NREMSAL-----SLPCKKP-----FGMTALDMKTRQVAVQVPLGITVEDTGLIGKGLK 736
Qy      488 SPV-----LSTAGGVENGCT-DRYFPALSOETGETLMQARLATVATGAISTY--EL 536
Db      737 APIGMPTIGGMAQGLVFFPAATQDYLLRAFDSTGKELKSMMPVGSQGTISVSPK 796
Qy      537 DGVOYIAGAGG 548
Db      797 TGKQYVVTAGG 808

```

```

RESULT 15
US-09-489-039A-12436
; Sequence 12436, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709,2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12436
; LENGTH: 602
; TYPE: PRY
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12436

```

Query Match 9.8%; Score 302; DB 4; Length 602;  
 Best Local Similarity 23.3%; Pred. No. 2,4e-19;  
 Matches 147; Conservative 81; Mismatches 214; Indels 190; Gaps 30;

```

Qy      81 QVPMHIDGVMYLANPBDVQALDAQTDGLIMEHRQLPRAVATLNAQDRKRGVALYGTIS 140
Db      23 EVTPIKVGDTLFLCTAHQRLFALDAATGKEKMHFPDQLNDPSF-----QHTCRGVS 75
Qy      141 LYSSWDN-----HLIALDMETGVVFPVERGSGE----- 170
Db      76 YHEKADNAPADVADDCRRRIILPVNDR-LFAVNADNGKLCETPANKGILNQTMPVT 134
Qy      171 -DGLTSNTGPIYAN-GVIYAGS-TCQYS---PYGCGISGDSATGEBELRNHFTIQ--- 221
Db      135 TPQNYEPTSPFIITDKTIVAGAVTDFSTREPSG--VIRGFVNTGKTLMA--FDPGARD 191
Qy      222 -----PGESEDETGNDEAFARMWTGVMGQITYPVTLNLYFYSTGVPAASETOGFGT 277

```

```

Db      192 PNAISDEHHTLINS-----PNSMAPAAYDAKLDLVYL-----PMGVTTPIDWGNR 238
Qy      278 YGNTRF-----AVRPDGEIYVRHQTLPRDNMQOECTFEMKVANVDVQSAEVEGLRALI 332
Db      239 TPEGERYASIVYALNATTKLMASTYQIVHNDLMDNDMSQPTLADIEV----- 286
Qy      333 NPMAATGERRVLTGAPCKTGTMTSFDPAASGEFL-----WARDTNY--TMTIASI 379
Db      287 -----NGKTVPVIVAPAKTGNIFVLDNRNGELVPAPEKVPQGAAGDYVTKQPFSDL 341
Qy      380 -----DETGL--VTNEDAVLKELDVEYD-----VCPPTFLGGRDMSA 415
Db      342 SFPKPKDLTGADMGATMFDQLVCSRVIHQLEKYEIGITPPSEQGLVFPFGNIGMEWGIGI 401
Qy      416 ALNPDTGIYF-----LPANNACY-----DINAVDOESALDVNTSATKLAP 458
Db      402 SYDPRQVAVANPMPALPFSKLIIRGPNPMPPKDAKSGTESGVQYGVPIYVTLNP 461
Qy      459 GFEEM-----GRIDAIDISTGRTLS-----AERPAANYSPVL--- 491
Db      462 FLSPPGLPCKOPAWGYISALDKTNEVWKKRIGTPQDSLFPMPVKLPFTMGEMTLGP 521
Qy      492 -STAGGVENGCT-DRYFPALSOETGETLMQARLATVATGA--ISYELDGVOYIAGAG 547
Db      522 ISTAGNVLEFIGATADNYLRAVYMSNGEKLWEARLP--AGGQATEPWTIEVNGQYVVISAG 579
Qy      548 GL-TYGTQLNAPLAEALIDSTSVGNALIVYFALP 578
Db      580 GHGSFGTKM-----GDYIVAVALP 598

```

Search completed: November 18, 2004, 05:21:08  
 Job time : 45 secs

*This Page Blank (uspto)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2004, 05:13:29 ; Search time 145 Seconds  
(without alignments)  
1414.068 Million cell updates/sec

Title: US-10-802-682-8  
Perfect score: 3069  
Sequence: 1 NMPTLLRTSAVLLTTPA.....AENIDSTGVNATVFPQ 579

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1570615 segs, 354127592 residues

Total number of hits satisfying chosen parameters: 1570615

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2765	90.1	580	14 US-10-162-713-8	Sequence 8, Appl
2	2569.5	83.7	578	14 US-10-162-713-5	Sequence 5, Appl
3	2559.5	83.4	579	14 US-10-162-713-7	Sequence 7, Appl
4	755.5	24.6	717	14 US-10-369-493-10840	Sequence 10840, A
5	624.5	20.3	567	14 US-10-369-493-7364	Sequence 7364, Ap
6	622	20.3	577	14 US-10-369-493-21739	Sequence 21739, A
7	620	20.2	692	14 US-10-369-493-20524	Sequence 20524, A
8	619	20.2	559	14 US-10-369-493-4607	Sequence 4607, Ap
9	578	18.8	623	14 US-10-369-493-8046	Sequence 8046, Ap
10	554.5	18.1	601	14 US-10-369-493-4930	Sequence 4930, Ap
11	547	17.8	592	14 US-10-369-493-7668	Sequence 7668, Ap
12	529.5	17.3	532	14 US-10-369-493-7687	Sequence 7687, Ap
13	524.5	17.1	706	14 US-10-369-493-10838	Sequence 10838, A

14	513.5	16.7	683	14 US-10-369-493-4929	Sequence 4929, Ap
15	482.5	15.7	592	14 US-10-369-493-15969	Sequence 15969, A
16	482.5	15.7	592	14 US-10-369-493-16311	Sequence 16311, A
17	482.5	15.7	593	14 US-10-369-493-15593	Sequence 15593, A
18	416	13.6	792	14 US-10-369-493-4878	Sequence 4878, Ap
19	412	13.4	803	10 US-10-369-493-7637	Sequence 7637, Ap
20	408	13.3	804	10 US-09-927-827-69	Sequence 69, Appl
21	408	13.3	804	14 US-10-369-493-4821	Sequence 4821, Ap
22	408	13.3	804	14 US-10-369-493-7580	Sequence 7580, Ap
23	407	13.3	798	14 US-10-369-493-15413	Sequence 15413, A
24	407	13.3	798	14 US-10-369-493-15781	Sequence 15781, A
25	407	13.3	798	14 US-10-369-493-16183	Sequence 16183, A
26	399	13.0	796	14 US-10-369-493-703	Sequence 703, Appl
27	373.5	12.2	777	14 US-10-369-493-8038	Sequence 8038, Ap
28	367.5	12.0	778	14 US-10-369-493-15182	Sequence 15182, A
29	363	11.8	786	14 US-10-369-493-16335	Sequence 16335, A
30	363	11.8	791	14 US-10-369-493-15596	Sequence 15596, A
31	363	11.8	802	14 US-10-369-493-15973	Sequence 15973, A
32	355	11.6	785	14 US-10-369-493-11883	Sequence 11883, A
33	354	11.5	783	14 US-10-369-493-14105	Sequence 14105, A
34	351.5	11.5	800	14 US-10-369-493-14064	Sequence 14064, A
35	342.5	11.2	783	14 US-10-369-493-11386	Sequence 11386, A
36	342.5	11.2	783	14 US-10-369-493-14430	Sequence 14430, A
37	342.5	11.2	783	14 US-10-369-493-14704	Sequence 14704, A
38	342.5	11.2	783	14 US-10-369-493-15187	Sequence 15187, A
39	324.5	10.6	788	14 US-10-369-493-11652	Sequence 11652, A
40	324.5	10.6	788	14 US-10-369-493-14659	Sequence 14659, A
41	323	10.5	787	14 US-10-369-493-14374	Sequence 14374, A
42	311	10.1	790	14 US-10-369-493-16392	Sequence 16392, A
43	176.5	5.8	564	14 US-10-369-493-17904	Sequence 17904, A
44	175.5	5.7	381	15 US-10-282-122A-50141	Sequence 50141, A
45	157.5	5.1	355	15 US-10-282-122A-69695	Sequence 69695, A

## ALIGNMENTS

RESULT 1  
US-10-162-713-8  
; Sequence 8, Application US/10162713  
; Publication No. US20030228672A1  
; GENERAL INFORMATION:  
; APPLICANT: Choi, Eun-Sung  
; APPLICANT: D'Elia, John  
; APPLICANT: Kim, Hye-Sun  
; APPLICANT: Kim, Xi-Soo  
; APPLICANT: Lee, Jung Kee  
; APPLICANT: Pan, Jae-Gu  
; APPLICANT: Stoddard, Steven F.  
; APPLICANT: Yum, Do-Young  
; TITLE OF INVENTION: Sorbital Dehydrogenases of Ketogulonigenium spp., Genes and Methoc  
; FILE REFERENCE: Use thereof  
; FILE REFERENCE: 1533.3740000  
; CURRENT APPLICATION NUMBER: US/10/162.713  
; CURRENT FILING DATE: 2002-06-06  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 580  
; TYPE: PRT  
; ORGANISM: Ketogulonigenium sp.  
US-10-162-713-8  
Query Match 90.1%; Score 2765; DB 14; Length 580;  
Best Local Similarity 87.6%; Pred. No 1e-236;  
Matches 507; Conservative 38; Mismatches 34; Indels 0; Gaps 0;  
QY 1 NMPTLLRTSAVLLTTPAFAQVPTITDELLANPAGGWINYGRQENTRHSPLTQIT 60  
DB 1 NMPTLLRTSAVLLTTPAFAQVPTITDELLANPAGGWINYGRQENTRHSPLTQIT 60  
QY 61 ADNVOGLDVAWAGAGAGVGTPTMHDGWNVYLANPGVIGALDAQGTDLTWEHRQCPA 120

Db 61 TDVVGQLQVWARGMEAGAVQVT PMIHGVMYLANGDVIAQIDAKTGDLMWEHRQLP 120  
Qy 121 VATLNAOGDRKRGVALYGTSLYFSSMDNLIALDMETGOVVPVVERSGSDGLTSNTGP 180  
Db 121 VASLNGDRKRGVALYGTSLYFSSMDNLIALDMETGOVVPVVERSGSDGLTSNTGP 180  
Qy 181 IVANGVIVAGSTCOYSPYGCFTSGHDSATGEBELMRNFIPOGSEDEFTMGNDPEARMT 240  
Db 181 IVANGVIVAGSTCOYSPYGCFTSGHDSATGEBELMRNFIPOGSEDEFTMGNDPEARMT 240  
Qy 241 GVMGQITPDTNLYVFGSTGVGPASSETGTGPGTLYGTNTREFAVRPDTGEIWMRHQTL 300  
Db 241 GVMGQITPDTNLYVFGSTGVGPASSETGTGPGTLYGTNTREFAVRPDTGEIWMRHQTL 300  
Qy 301 PRDWMDECTEFEMVAVADVQPSAEMEGRLAINPNAATGERRVLITGAPCKTGTMWSDAA 360  
Db 301 PRDWMDECTEFEMVAVADVQPSAEMEGRLAINPNAATGERRVLITGAPCKTGTMWSDAA 360  
Qy 361 SGEFLMARDNTYNTMISIDETGLVTNEDAVLKEIDVEYDVCPTFLGSRDWSAALNDP 420  
Db 361 SGEFLMARDNTYNTMISIDETGLVTNEDAVLKEIDVEYDVCPTFLGSRDWSAALNDP 420  
Qy 421 TGIYFLPLNNAACVDINAVDOEFSAIDVYNTSATAKLAPGEMNGRIDAIDISTGRTLMSA 480  
Db 421 TGIYFLPLNNAACVDINAVDOEFSAIDVYNTSATAKLAPGEMNGRIDAIDISTGRTLMSA 480  
Qy 481 ERPAANSPLYSTAGGVFNGGTRFYRALSOETGELTMOARLATVATGQALSYELDGVQ 540  
Db 481 ERPAANSPLYSTAGGVFNGGTRFYRALSOETGELTMOARLATVATGQALSYELDGVQ 540  
Qy 541 YVIAAGGLTYGTQLNAPLAEIDSTSVGNATVYFALPQ 579  
Db 541 YVIAAGGLTYGTQLNAPLAEIDSTSVGNATVYFALPQ 579

RESULT 2  
US-10-162-713-5

; Sequence 5, Application US/10162713  
; Publication No. US20030228672A1  
; GENERAL INFORMATION:  
; APPLICANT: Choi, Eun-Sung  
; APPLICANT: D'Elia, John  
; APPLICANT: Kim, Hye-Sun  
; APPLICANT: Kim, Mi-Soo  
; APPLICANT: Lee, Jung Kee  
; APPLICANT: Pan, Jae-Gu  
; APPLICANT: Stoddard, Steven F.  
; TITLE OF INVENTION: Sorbitol Dehydrogenases of Ketogulonigenium spp., Genes and Method  
; FILE REFERENCE: 1533.3740000  
; CURRENT APPLICATION NUMBER: US/10/162,713  
; CURRENT FILING DATE: 2002-06-06  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 578  
; TYPE: PRP  
; ORGANISM: Ketogulonigenium sp.  
US-10-162-713-5

Query Match 83.7%; Score 2569.5; DB 14; Length 578;  
Best Local Similarity 80.1%; Pred. No. 2,4e-219;  
Matches 466; Conservative 59; Mismatches 55; Indels 1; Gaps 1;  
Qy 1 MNPTTLRTSAVLLTAPAAFAQVTPITDELANPPAGEMINYGRCNENYRHSPLTQT 60  
Db 1 MKNSLILSAVAVAFPAFADVTPVDELANPPAGEMISYGRNQENYRHSPLNQIT 60  
Qy 61 ADVNGQLQVWARGMEAGAVQVT PMIHGVMYLANGDVIAQIDAKTGDLMWEHRQLP 120  
Db 61 ADVNGQLQVWARGMEAGAVQVT PMIHGVMYLANGDVIAQIDAKTGDLMWEHRQLP 120  
Qy 121 VASLNGDRKRGVALYGTSLYFSSMDNLIALDMETGOVVPVVERSGSDGLTSNTGP 180  
Db 121 VASLNGDRKRGVALYGTSLYFSSMDNLIALDMETGOVVPVVERSGSDGLTSNTGP 180

Qy 121 VATLNAOGDRKRGVALYGTSLYFSSMDNLIALDMETGOVVPVVERSGSDGLTSNTGP 180  
Db 121 TSTLSSLDGRKRGVALYGTNTVYFVSNMNVVALDAAAGQVVFVDRGQSDERV-SNUSGP 179  
Qy 181 IVANGVIVAGSTCOYSPYGCFTSGHDSATGEBELMRNFIPOGSEDEFTMGNDPEARMT 240  
Db 181 IVANGVIVAGSTCOYSPYGCFTSGHDSATGEBELMRNFIPOGSEDEFTMGNDPEARMT 240  
Qy 241 GVMGQITPDTNLYVFGSTGVGPASSETGTGPGTLYGTNTREFAVRPDTGEIWMRHQTL 300  
Db 241 GVMGQITPDTNLYVFGSTGVGPASSETGTGPGTLYGTNTREFAVRPDTGEIWMRHQTL 300  
Qy 301 PRDWMDECTEFEMVAVADVQPSAEMEGRLAINPNAATGERRVLITGAPCKTGTMWSDAA 360  
Db 301 PRDWMDECTEFEMVAVADVQPSAEMEGRLAINPNAATGERRVLITGAPCKTGTMWSDAA 360  
Qy 361 SGEFLMARDNTYNTMISIDETGLVTNEDAVLKEIDVEYDVCPTFLGSRDWSAALNDP 420  
Db 361 SGEFLMARDNTYNTMISIDETGLVTNEDAVLKEIDVEYDVCPTFLGSRDWSAALNDP 420  
Qy 421 TGIYFLPLNNAACVDINAVDOEFSAIDVYNTSATAKLAPGEMNGRIDAIDISTGRTLMSA 480  
Db 421 TGIYFLPLNNAACVDINAVDOEFSAIDVYNTSATAKLAPGEMNGRIDAIDISTGRTLMSA 480  
Qy 481 ERPAANSPLYSTAGGVFNGGTRFYRALSOETGELTMOARLATVATGQALSYELDGVQ 540  
Db 481 ERPAANSPLYSTAGGVFNGGTRFYRALSOETGELTMOARLATVATGQALSYELDGVQ 540  
Qy 541 YVIAAGGLTYGTQLNAPLAEIDSTSVGNATVYFALPQ 579  
Db 541 YVIAAGGLTYGTQLNAPLAEIDSTSVGNATVYFALPQ 579

RESULT 3  
US-10-162-713-7

; Sequence 7, Application US/10162713  
; Publication No. US20030228672A1  
; GENERAL INFORMATION:  
; APPLICANT: Choi, Eun-Sung  
; APPLICANT: D'Elia, John  
; APPLICANT: Kim, Hye-Sun  
; APPLICANT: Kim, Mi-Soo  
; APPLICANT: Lee, Jung Kee  
; APPLICANT: Pan, Jae-Gu  
; APPLICANT: Stoddard, Steven F.  
; TITLE OF INVENTION: Sorbitol Dehydrogenases of Ketogulonigenium spp., Genes and Method  
; FILE REFERENCE: 1533.3740000  
; CURRENT APPLICATION NUMBER: US/10/162,713  
; CURRENT FILING DATE: 2002-06-06  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 7  
; LENGTH: 579  
; TYPE: PRP  
; ORGANISM: Ketogulonigenium sp.  
US-10-162-713-7

Query Match 83.4%; Score 2559.5; DB 14; Length 579;  
Best Local Similarity 80.5%; Pred. No. 1.9e-218;  
Matches 466; Conservative 52; Mismatches 60; Indels 1; Gaps 1;  
Qy 1 MNPTTLRTSAVLLTAPAAFAQVTPITDELANPPAGEMINYGRCNENYRHSPLTQT 60  
Db 1 MKTSFLFAGVAAALASGTIALADVTPVDELANPPAGEMISYGRNQENYRHSPLNQIT 60  
Qy 61 ADVNGQLQVWARGMEAGAVQVT PMIHGVMYLANGDVIAQIDAKTGDLMWEHRQLP 120  
Db 61 ADVNGQLQVWARGMEAGAVQVT PMIHGVMYLANGDVIAQIDAKTGDLMWEHRQLP 120  
Qy 121 VASLNGDRKRGVALYGTSLYFSSMDNLIALDMETGOVVPVVERSGSDGLTSNTGP 180  
Db 121 VASLNGDRKRGVALYGTSLYFSSMDNLIALDMETGOVVPVVERSGSDGLTSNTGP 180

```

Db 121 VALLNSGEBIRGIALYNTNVFVSWDNHLVALDAATGQVTFPDVDEGQED-MVNSSGP 179
Qy 181 IVANGVIVASTCOQSPYSGCFISGHDSATGSELNRNHFIPQGEEDGETWGNDFEARMT 240
Db 180 IVANGVIVASTCOQSPYSGCFISGHDSATGSELNRNHFIPQGEEDGETWGNDFEARMT 239
Qy 241 GWMQITVDPVTVLVPYSGTGVGPASSETQGTGPTGTLVGNTRFAVRPDTGEIWMRHQTL 300
Db 240 GWMQITVDPVTVLVPYSGTGVGPASSETQGTGPTGTLVGNTRFAVRPDTGEIWMRHQTL 299
Qy 301 PROWDQECTFEMWVAVADVOPSAEMEGRLAINPNAATGERRLVJGAPCKGTWMSFDDA 360
Db 300 PROWDQECTFEMWVAVADVOPSAEMEGRLAINPNAATGERRLVJGAPCKGTWMSFDDA 359
Qy 361 SGEFLMARDNTYNTMLASIDETGLVTVNEBDAVLKELDEVEDVCPTEFLGGRDWSAALNDP 420
Db 360 TGEFLMARDNTYNTMLASIDETGLVTVNEBDAVLKELDEVEDVCPTEFLGGRDWSAALNDP 419
Qy 421 TGIYFLPNNACVDIMAVDOEFSLDVTNTSATAKLAPGEMNGRIDALDISGRTLMGA 480
Db 420 SGIYFLPNNACVDIMAVDOEFSLDVTNTSATAKLAPGEMNGRIDALDISGRTLMGA 479
Qy 481 ERPAANSPIVSTAGGVFNQGTDRYFPAISOETGETLMQARLATVATGOAISYEIDGVQ 540
Db 480 ERPAANSPIVSTAGGVFNQGTDRYFPAISOETGETLMQARLATVATGOAISYEIDGVQ 539
Qy 541 YIAIAGGLTYGTQLNAPLAELDSTSVGNATVYFALPQ 579
Db 540 YIAIAGGLTYGTQLNAPLAELDSTSVGNATVYFALPQ 578

```

RESULT 4

```

US-10-369-493-10840
; Sequence 10840, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10840
; LENGTH: 717
; TYPE: PRT
; ORGANISM: Sphingomonas aromaticivorans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(717)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-10840

```

```

Query Match 24.6%; Score 755.5; DB 14; Length 717;
Best Local Similarity 33.6%; Pred. No. 7,7e-58;
Matches 187; Conservative 90; Mismatches 217; Indels 63; Gaps 17;

Qy 28 ITBELANPPAGEINYGRENQENYRHSPLQITADNVGOLQV-----ARMEAGAVQ 81
Db 38 VTDALLIOAREGEMLSGRDYGEORSPLOINDANGVGLMFPDLERARQEA----- 93
Qy 82 VTPMINDGVYLANPGDVIQALDAQTGDLIMEHRROLPAVALTNAQDR-KRGVALYGTG 140
Db 94 -TFLMHDGTLVYSTAWGMKAFDAKGTALKMSYDPEVRETLVACDVAWGSVALYDK 152
Qy 141 LYSSSWNHLLALDMEGGVVFDVERSGDGLTSNTGTIVANGVIVASTCOQSPYGC 200

```

```

Db 153 VEVGLDRLVALDQKTKGVWVKVVPNOEDYITTGAPRVVKGKVLIGSGGSSEYKAG- 211
Qy 201 FTSCHDSATGSELNRNHFIPQGEEDGETWGNDFEARMTG-----WMOGIT 247
Db 212 YIAADVNTGNEVKKFHFVPGNPDGPFENKAMENNAARWAGEW---WELGGGGTYMDSIT 268
Qy 248 YDPVTVNLVYSGTGVGPASSETQGTGPTGTLVGNTRFAVRPDTGEIWMRHQTLPRDMDQ 307
Db 269 YDPVTVNLVYSGTGVGPASSETQGTGPTGTLVGNTRFAVRPDTGEIWMRHQTLPRDMDQ 327
Qy 308 ECTEFMVAADVOPSAEMEGRLAINPNAATGERRLVJGAPCKGTWMSFDDAASGEFLW 366
Db 308 DSAQOITLADLTID-----GQRHVLHAP-KXGHVYVLDAARTGOFLS 369
Qy 367 ARDTYNTMIAISID-ETGLVTVNEBDAVLKELDEVEDVCPTEFLGGRDWSAALNDPDTGYF 425
Db 370 ATPFVVMWATGIDPKTKATVNPBARYEKTKGFVSLPGAVGAHSMQPFSPKGTGLY 429
Qy 426 LPNNACT-----DINAVDOEF-SALDVNTS--ATKLT--AGFEMNGRIDALDISG 474
Db 420 LPNNAAFPYAAAKCWKATDIDGFQGLDGYVTSMPADAKVGAAMKATTTGLVWMDPVAK 489
Qy 475 RTLWGAERPAANSPIVSTAGGVFNQGTDRYFPAISOETGETLMQARLATVATGOAISY 534
Db 490 KRAMVVELPSBNSNGSITLTAQNLVYGTAGDPAVAYNADKQKQMLPFAOSGIIAAPTLY 549
Qy 535 EIDGVQYTA--IGAGGL 549
Db 550 AIDGEOYAVWVGWGV 566

```

RESULT 5

```

US-10-369-493-7364
; Sequence 7364, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 7364
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-369-493-7364

```

```

Query Match 20.3%; Score 624.5; DB 14; Length 567;
Best Local Similarity 28.5%; Pred. No. 2,4e-46;
Matches 172; Conservative 105; Mismatches 257; Indels 69; Gaps 21;

Qy 3 PTTL-LRTSAVLLITAPAFAPQVTPIT--DELIANPPAGEINYGRENQENYRHSPLQI 59
Db 6 PTALMSVAVAVVSVLFSGSAAADYPAVYERLTAQSDPMLTYRYTNGOASPLKQI 65
Qy 60 TADNVGOLQVLA-----RGMAGAVQVTPMINDGVYLANPGDVIQALDAQTGDLI 111
Db 66 DTANVAKLKQWMSYKFPADLDQGFBA-----TPIVNGYLVLTTPKXNVVAFDAATGQL 120
Qy 112 WEHRRLDPAVALTNAQGD-RRGVALYGTSLYFSSWMDNHLLALDMETGQVFDVERSGGE 170
Db 121 WFEFKLAEBSFKACCBVIRGVALYKNTYVAMLGSDVVALDAQIGALAW--RKQWFE 178
Qy 171 DGL-TSNTTGTIVANGVIVAST--COQSPYSGCFISGHDSATGSELNRNHFIPQGEED 228

```

Db 179 PGLGVAFLAPLALDGLVVGAGGEGARG-FIALNPDNGNVLWKEFTVPAGEKAD 237  
QY 229 TWGNDEAR-----WMITGVWGQITVDPTNLVFGSTGVGPASETORGTGGTLYGTNR 283  
Db 238 TWNGNGEHCAGPAMLTG-----TYDASKTLVWGNGPMLADR--PGNLY-SDSL 289  
QY 284 FAVRPTGTGLVWHTQTLPRDNWDOECTFEFMMVAVNDVQSAEMEGRAINPNAATGERV 343  
Db 290 LADPRTGDKMNYOYTKHDITWDYDVNTPVLTATYYQ-DKEEDALIHARN----- 340  
QY 344 LTGAPCTGTWMSFDDAASGEFLWARDNTYNTMIASIDETGLVT---VNEDAVLKEJVE 399  
Db 341 -----GFFHALDGTGLYAKP-----FTATSVYCTPADGAPIDASKYPRAGTT 387  
QY 400 YDVCPFTLGRDMSAALNDPTGIYPLPLNNACYDIMAVDOEF-SALDYNTSATAKLAP 458  
Db 388 IETCPSTLGGKMWMSISYDEPKHIAVVPALHACMSISGKSVNMEGLPYLGGFELKPRP 447  
QY 459 GFENMGRIADIDISTGRTLMSARPAANYSPVLSTAGVVFNGGTDRTYFRALSOETGETL 518  
Db 448 GSKYGELOAINVDITGKWSHMSKLPWNGVATTGGLAFSGSLDGHLYADPTTGKYL 507  
QY 519 WCA-RLATVATGQAISELDGVYIAGAGLTYGTQ---LNAFLAIDSTSVGNALYV 574  
Db 508 WQBPKLASGIVAQPSYFEVVDGKEYVALAG---YGANPIWGGPMAKAEKVPBGGLTV 564  
QY 575 PAL 577  
Db 565 PAL 567

## RESULT 6

US-10-369-493-21739  
; Sequence 21739, Application US/10369493  
; Publication No. US2003023675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 21739  
; LENGTH: 577  
; TYPE: PRT  
; ORGANISM: Rhodobacter capsulatus  
US-10-369-493-21739

Query Match 20.3%; Score 622; DB 14; Length 577;  
Best Local Similarity 29.6%; Pred. No. 4.2e-46;  
Matches 186; Conservative 98; Mismatches 231; Indels 114; Gaps 28;

QY 11 AAVLLTPAPAFQVPTIDELAN--PPAGEWINGRNOENYRHSPLTOITADNQGOL 68  
Db 2 AAVALTLC-FGTADAGVTEDLRNDQGTAGDVNTGMDLDRFPLATLNKRDVKNLM 60  
QY 69 LWNA-----RGEAGAVQVTPMTHDGVWYLANPBGVIALDAOTGLIMEHRQLPA 120  
Db 61 PAAAFSLGSEKRGQES-----OPIVDGVMTITGSYSRLYALDITGKEIMWOYDARLE 115  
QY 121 VATLNAQDGRKRGVALVGTSLYSSWMDNLIALDMETGVVFPVERSGEGDGLTSNTGP 180  
Db 116 -GILPCDVTNRGAALIGDGVYFTGLDARIVALDRKTGVKMKKADYKEGI-STTAAP 173  
QY 181 IVANGVIVAGSTQCYSPYGC--ISGDSATGEELWNRHFIPOPGEG----- 226

Db 174 LTVNGLIVTGN--GSEFGLVGEVQAKNAETGELWTRPMI--EGHMTLNGKPMTGT 229  
QY 227 -DETGNDFEARWMTG---VWGQITVDPTNLVFGSTGVGPASETOR--GTP-----GG 275  
Db 230 LNAITWPDGL---WKYGGATWLGSSYDAETDLTVFGAGNPAPNNSHLRNAGTPEGNKGD 286  
QY 276 TLVGTNRFRANRPTGTGLVWRHDTLPRDNWDOECTFEFMMVAVNDVQSAEMEGRAINPN 335  
Db 287 NLVRA-SRVDGPTGLKMHFQITTPREGWDFGVNE-VGFPD-----KDG---NKR 334  
QY 336 AATGEREVLTPAPCKTGTWMSFPAASGEFL--WARDNTYNTMIASIDETGLVTNED-- 390  
Db 335 FATADR-----NGFYVLNRADGKFVAMPFVGNIT--WAKGIDETGRPIYNEHRNP 384  
QY 391 -----AVLKELDYDVCPFTLGRDMSAALNDPTGIYPLPLNNACYDIMAVDOERSAL 445  
Db 385 GAPPAALAAKQGITFAVPSFTLGGKMWMSISYDEPKHIAVVPALHACMSISGKSVNMEGLPYLGGFELKPRP 442  
QY 446 DVYN-----TSATAKLAPF-----ENMGRIADIDISTGRTLMSARPAANYSPYLST 493  
Db 433 DINNEPINKKGAAYLAGFTIKLPEDIYGLKAIIDPNTGAVKMDAPLMGVMTT 492  
QY 494 AGSVFNGGTDRTYFRALSOETGETLQARLATAVATGQAISELDGVYI--GAGLT- 550  
Db 493 AGKLVFTGNPEGEFWALDAETGKLMSPQTSGSIVGQPIWEDDGEQVSIISGWSGAVP 552  
QY 551 -YGTQNLAPLAIDSTSVGNALYVFPAL 578  
Db 553 LMGE-----VAKRNVYLNQGTWTFRLP 577

## RESULT 7

US-10-369-493-20524  
; Sequence 20524, Application US/10369493  
; Publication No. US2003023675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 20524  
; LENGTH: 692  
; TYPE: PRT  
; ORGANISM: Rhodospseudomonas palustris  
US-10-369-493-20524

Query Match 20.2%; Score 620; DB 14; Length 692;  
Best Local Similarity 29.3%; Pred. No. 8.3e-46;  
Matches 180; Conservative 100; Mismatches 246; Indels 88; Gaps 23;

QY 15 LITAPAAFAQ-----VTPITDE--LLAN-PPAGEWINGRNOENYRHSPLTOIT 60  
Db 5 LITPAVAAQDSKSAHIRAVTGAVDSAAIVAVKTKTQMPSGLDYAFTRFSKIDQIN 64  
QY 61 ADNVGQLOLV--ARGMEAGAVQVTPMTHDGVWYLANPBGVIALDAOTGLIMEHRQL 118  
Db 65 TENKQGLQWSYSLGSERG--VEATPVAVDGMIVTASWGVVAHIDRTGKLTPEPKV 123  
QY 119 PAVATLNAQDR-----KRGVALYGTSLYSSWMDNLIALDMETGVVFPVERSGEGD 172  
Db 124 D-----HSKYGRCGCDVNRGVALYKGVFVGAVDGLIALDAATGSKAMEIDTLIDHEH 178  
QY 173 LTSNTGTIVANGVIVAGS--TCQYSPYGCFLISGDSATGEELWNRHFIPOPGEG-----OP 222

Db 179 SYITGAPRVNGKVIIGNGAEYGAR-G-VTAAYDAETGKQAMRFTVPDPSKEPDES 237  
 QY 223 GEBDETFGNDPEFARWMTG----WGQITYPVTNLVFGSTGVPASETORGTGGTLY 278  
 Db 238 MEKRAKTM-DPAKMWLNGGGGTAMOTITFPDPLNLIVGTGNGSPMNHLSRPAAGDNL 296  
 QY 279 GTTTRFVRPDGTGEIYWRHQTLPREDWDECTEFEMVAN--VDVQSAEMEGRLAIPNA 336  
 Db 297 YLASIYALNADTKYVWHYQETPGDWDTYSTQPMILADIAIDKP----- 342  
 QY 337 ATGERRLTGAPCKGTGTMGTFDASGEFLWADNTNTNMIASIDEGL-VTVNEDAVLKE 395  
 Db 343 ----RKIYLHAP-KNGFFVIDRTDGFISAKNFVDVMAATGDSNGRPEVE---AS 394  
 QY 396 LDVEYDVCPFTFGDWSAALNPDGTGYFLPLNNAACYDI--MAVDQ-----EFSALD 446  
 Db 395 ADSKFDALPGPYGAHMHHPMSFNPQTGLVYLPAGQVAVNLGKALTONMEPEFKGSGTT 454  
 QY 447 VYNTSATKLAAPGEMN--GRIDAIDISTGRTLSAERPAANYSPVLTAGVVFNGGTD 504  
 Db 455 GNVVGFLLNAVPP-KNLPFGELVAMDPVQCKEYWRLEIYSPMNGTLLTRAGNLVFGCTAD 513  
 QY 505 RYFRALSOETGELTMCARLATVATGCAISYELDGVQYIAIGAG-GLTGTOLNAPLAEX 563  
 Db 514 GRFVAVYAKTGEKLMESPLGTGAVAAPATYWDGVQYVIAVGMGVFGISARATETEA- 572  
 QY 564 DSTSVGNATVFPAL 577  
 Db 573 ----PGTYVTFAN 581

## RESULT 8

US-10-369-493-4607  
 ; Sequence 4607, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; PRIOR FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 4607  
 ; LENGTH: 559  
 ; TYPE: PRT  
 ; ORGANISM: Burkholderia fungorum  
 US-10-369-493-4607

Query Match 20.2%; Score 619; DB 14; Length 559;  
 Best Local Similarity 28.3%; Pred. No. 7.4e-46;  
 Matches 169; Conservative 105; Mismatches 256; Indels 68; Gaps 20;

QY 7 LRTSAALLTLTAPAFQVTPIT--DELLANPRAGEWINGRNOENYRHSPLQITADNY 64  
 Db 3 MSVAVASLVFGSAALADYPAVYERLTAQSDPGMLTYRYYNGOAHSPLOITANV 62  
 QY 65 GQDOLVNA-----RGEAGAVQVTPMTHDGVMYLANPDVIALDAQDGLIWEHNR 116  
 Db 63 KNLQVMSYKRPADLQGGFEA-----TPIVNGRYLFTVTPKDNVYAFDATTGQWTFEP 117  
 QY 117 QLPRAVATLNAQD-RKRGVALYGTSLYFSSMDNHLILADMETGQVVDVERSGEGCL-T 174  
 Db 118 KLGESSEKTAQCDVINGVALYGGKVVYAMLSGDVVALDQGTGLAW--RKQFFRGLGY 175  
 QY 175 SNTTGPVANGVAVAGST-CQSPYGCFIGSHDSATGEELMRNHFLFQPGEGSDETWGD 233

Db 176 AFSLAPLADGALVVGAGGEYGAR-G-FIALANDNGNVLKRTVPAGKGDTPNG 234  
 QY 234 FEAR-----NMWGWGQITYPVTNLVFGSTGVPASETORGTGGTLYGNTFRFAVR 288  
 Db 235 MQEHGADPAMLTG-----TYDAASKTLYWGVNGPMLADR--FGDNLV-SDSILALDF 286  
 QY 289 DTGEIYWHQQLPREDNMQECTEFEMVANVDVQSAEMEGRLAIPNAATGERVLTGAP 348  
 Db 287 KTGDKMHVQYTKHDPTMDYDGVNTPVLTIRYQ-DKEVDALIHADR----- 332  
 QY 349 CKTGTMGTFDASGEFLWADNTNTNMIASIDEGLVTV--VNEDAVLKELDVEYVCP 404  
 Db 333 ---GFFHAILDGTGLIYAKP-----FVLATSVGYADAGAPIDQASKYPRAGTTIECP 384  
 QY 405 TFLGGRDWSAALNPDGTGYFLPLNNAACYDI--MAVDQ-----EFSALD 463  
 Db 385 SFLGGRDWSAALNPDGTGYFLPLNNAACYDI--MAVDQ-----EFSALD 463  
 QY 464 GRIDAIDISTGRTLSAERPAANYSPVLTAGVVFNGGTD RYFRALSOETGELTMOA-R 522  
 Db 445 GELQALINDTKRWKWSHMSKLPNMGVATTGGLAFSGSLDGLIYAPDETGTGKLMQSPK 504  
 QY 523 LATVATGCAISYELDGVQYIAIGAGLTGTQ---LNAPLAEXDSTSVGNATVFPAL 577  
 Db 505 LASGIVAPSVFEVDGKEYVAIAG---YGGANPIWGGPMKAAEKYPRGGTLYVFPAL 559

## RESULT 9

US-10-369-493-8046  
 ; Sequence 8046, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; PRIOR FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 8046  
 ; LENGTH: 623  
 ; TYPE: PRT  
 ; ORGANISM: Rhodobacter sphaeroides  
 ; NAME/KEY: unsure  
 ; LOCATION: (1) (623)  
 ; OTHER INFORMATION: unsure at all Xaa locations  
 US-10-369-493-8046

Query Match 18.8%; Score 578; DB 14; Length 623;  
 Best Local Similarity 27.8%; Pred. No. 3.8e-42;  
 Matches 177; Conservative 103; Mismatches 259; Indels 98; Gaps 22;

QY 3 PTLTSTS-AAVLTLTAPAFQVTPIT--DELLANPRAGEWINGRNOENYRHSPLQIT 60  
 Db 10 PMKMLKTGVALTLLSSAPVFA-----NDVYLKATDDPKQALQCGDVANRYSELNGIT 64  
 QY 61 ADVGQDOLVNA-----RGEAGAVQVTPMTHDGVMYLANPDVIALDAQDGLIWEHNR 112  
 Db 65 RENVEKLGQVAMFTSTGVLRGHEG-----SPVVGIMVYVHTPPNNVVALDLANDGKLIW 119  
 QY 113 EHR-ROLPAVATLNAQDGRKRGVALYGTSLYFSSMDNHLILADMETGQVVDVERSGEGCL 171  
 Db 120 RYEPQNDPVIGVCMCCDTVNRGVALYADGMIFLHQADITVVALDPAKSGVKTIVQNGDPAK 179  
 QY 172 GLTSTTGPVANGVAVAGSTCQSPYGCFIGSHDSATGEELMRNHFLFQPGEGSDETWGD 219



Db 177 KVLTSIGSGEFGVGRLLA-YDIKTKGPAWTAISTGDKMLIDPKTTTYADGRVWPV 235  
 Qy 223 GEEGDETFWCHDFEARMWTG---VWGQITTYDPTNLVFGSTGVGPASETORGTGGTLXG 279  
 Db 236 ADSLSKWKBD---QWMLGGGTTMGWYAMPKLNLYVGTGNGTNNPTOR--FGNNKMS 290  
 Qy 280 TNRFAVRPTGTGLVWRHQTLPNDNDQECTFEMMANVDVQ---PSA---EMEGRLAI 332  
 Db 291 MSI-FARDLNTGQARVWYQMTPHDEWDYDGVNEMILSDLSIDKQVPAIVHPRNGF-GY 348  
 Qy 333 NPNATGEREVLTCAPCKTGTMMS--FDAASGEFLMARDNTNNMTASIDETGLVTVNED 390  
 Db 349 TLRETFQ--LVAQKDPRAVNMADHYDMKSGPI--RNAVSTQAGSHN----- 396  
 Qy 391 AVLKEIDVEYDVCPTFLGHDWSSALNPDGTGYFLPLNACYDIMAVDQESALDVNT 450  
 Db 397 -----VKGICPAALGSKDQCPAAYDPSSSLFLVPTNHVCMDEPFDVYVSGQPY-V 447  
 Qy 451 SATAKLAPG---FENMRIDAIDISTRTLSAERPAWSPYLSAGVYFNGGTDYF 507  
 Db 448 GATLSMTPGNDNNSWGNFTAMWDSKGIWMSKPERPVSWSGLATAGVAFYGLTGYI 507  
 Qy 508 RALSOETGETLMQARLATAVATGQAI SYELDGVQYIAI--GAGGILTYGTOLNAPLAEIDS 565  
 Db 508 KAVRIKQKELMRKTPSGIIGNVFTYEQKQFIVYSGIGGWA-GIGWAAGLEKSTEG 566  
 Qy 566 -----TSVGNATVFPALP 578  
 Db 567 LGAVGYRELAKYATLGGTLFVFAIP 592

RESULT 12

US-10-369-493-7687  
 ; Sequence 7687, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xiandeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR FILING DATE: 2002-02-21  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 7687  
 ; LENGTH: 532  
 ; TYPE: PRT  
 ; ORGANISM: Burkholderia cepacia  
 ; US-10-369-493-7687

Query Match 17.3%; Score 529.5; DB 14; Length 532;  
 Best Local Similarity 27.6%; Pred. No. 6.3e-38;  
 Matches 161; Conservative 93; Mismatches 245; Indels 85; Gaps 20;

Qy 27 PINDELL--ANPAGEWYNGRQENYRHSPLQITDNNYGOQ--LVMAKGEAAVQV 82  
 Db 2 PVSQEQDRATTSNDLLNSGSAQTRYRGAQINNTVSKLPAPIFQTANF-SMET 60  
 Qy 83 TPMHDSVMTLANPGVYQALDAQTDGLIWEHRROLPAVATLNAQGRKRGVALYGTSLY 142  
 Db 61 APTSNVMTFTTSFNVAVADATGKEFWHYKHMGAVTF--CCGNRRGVAIADGRLY 119  
 Qy 143 FSSWDNHLIALDMETGGVPEVERGSGEDGITSITTPRIYANGVYAGSTCYSPGCC-- 200  
 Db 120 MGLTDALVALDATTGSLVWQTLADPDEG-SETVAPTVADKVLGT--NGEGEING 176  
 Qy 201 FISGHDATGSELNRNFIPOGEGEDETW-----GNDFEA 236

Db 177 FLKAFDANSQGLLMTFTYIIPETQEG--VMATKDATGDKRIDDAEKKQIAEKGGDFYK 234  
 Qy 237 RMTGVMQCIITYDPTNLVFGSTGVGPASETORGT--PGGTLVGTNTRFVRPDTGEIV 294  
 Db 235 TLGGVWVMAPIADRTHTVVF---VGNPSPDLYGAIKPGNLY-TDSLVAIDIDTGYK 290  
 Qy 295 WRHQTLPNDNDQECTFEMMANVDVQPSAMEGLRAINPNAAGGERVLTGAPKXTGM 354  
 Db 291 WHQVYPPHDVNDLDAVSPMLIDV-----RDNNGMIRQ---VHGG--KTGHV 334  
 Qy 365 WSFDAASGEFLMARDNTNNMTASIDETGLVTVNEDAVLKEIDVEYDVCPTFLGGRDSS 414  
 Db 335 YHDDATGRLI-----RISQ--AMTPOENMTTLPTAAGARPL-----PGANGVEMSP 380  
 Qy 415 AALNPDGTGYFLPLNACYDIMAVDQESALDVNTSATKLABFENMRIDAIDISTG 474  
 Db 381 NAFDPTRLVYAANLHQMTYQVEDAAYPGGSKMLGAFKTIASEQQMGKLSAVNDTG 440  
 Qy 475 RLMSAERPAWSPYLSAGVYFNGGTDYFRLASGETETLMQARLATAVATGQAI SY 534  
 Db 441 KVAADYKTEOPLIGSLVLTAGSLVFNGBNGLFRAPDSATKRLWEPCCGAVNAPAVSY 500  
 Qy 535 ELDDVQYTAIGAGSLTYGTOLNAPLAEIDS VSGNATVFPALP 578  
 Db 501 MVHGKQYIAVAAAG--NTQLD-----FKGNVTLVFPALP 532

RESULT 13

US-10-369-493-10838  
 ; Sequence 10838, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xiandeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR FILING DATE: 2002-02-21  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 10838  
 ; LENGTH: 706  
 ; TYPE: PRT  
 ; ORGANISM: Sphingomonas aromaticivorans  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(706)  
 ; OTHER INFORMATION: unsure at all Xaa locations  
 ; US-10-369-493-10838

Query Match 17.1%; Score 524.5; DB 14; Length 706;  
 Best Local Similarity 29.0%; Pred. No. 2.7e-37;  
 Matches 177; Conservative 79; Mismatches 244; Indels 111; Gaps 25;

Qy 15 LITAPAAQAVPTPINDELLANPAG--EWINNGRQENYRHSPLQITDNNYGOQOLV-- 71  
 Db 22 IASPPAPSAANVLE-----GAGREWPQGGDGKTHSRRLTRINANENVRGLGLAKOV 75  
 Qy 72 ---ARGNEAGAVQVTPMHDGVMTLANPGVYQALDAQTDGLIWEHRROLPAVATLNAQ 127  
 Db 76 ELGTLRQGEA---TPVVVGVLYTSGTGAAYAFDAATGKELMRFPEDVQVNRVVC 130  
 Qy 128 GGR-KGVYALGTSYFSSWDNHLIALDMETGOV---FVERGSGEDGILSTGTTG--PI 181  
 Db 131 CDVNRGVAVGKGVFVASLDGMMYALDARTGAVWWSDFLENKAGD---NSTGAP 185  
 Qy 182 VANGYIVAG-STCYSPYCGFISGHDATGSELNRNFIPOGEGED-----TW 230

Db 186 IAGDVVVGKAGBAYVRG-YTALDIDTGKLRNHHVPRDKLGQDETPELEALKTW 244

Qy 231 GNDPEARMWTG---VWGQITTDPTNLVFGSTGVGPASETGRTGGTLVGTNTRFV 286

Db 245 --DENSMDIGGGSPMDALINVDPEGLVAVGTGNGGPATSKRSPAGGDNILVLA 302

Qy 287 RPTDGLVEMHQTLPNNMQECTPEMMVANDVQSAEMEGRLAIPNNAATGHERVLTG 346

Db 303 DPKGRKRMHYQETPGNMDFTATQPMIFTRKIDG-----DRPVLAH 346

Qy 347 APCGTGMMGFDAASGEFLWARDNTYNNIASIDE-TGLVTNEDAVLKELDEYD---- 401

Db 347 AP-KNGVLTYLDRDGLLAAPIVRNNMAKGDPTG-----RILDPAADYTGRK 399

Qy 402 -VCEPFLGSDWSSALNPTGIY-----FLP-----LNNACDIYMDV 439

Db 400 IVEPATGAAHWHFASDPATGLYIGAVLDGMILFMTPGAKELKARGLNDDAALLFTPD 459

Qy 440 --GEFSALDVYNTSATUKLAPGEN-----NGRIDALDISTGRTLSAERPA-ANYSPV 490

Db 460 VKELAAFPPEPFGDAVKL-PAYQELAKOPATQAIIDPLTKITWADDTAGWQDRGV 518

Qy 491 LSTAGGVVNGGTDYFRALSOETGTLWQARLATVATGQALSYELDGYQYIAI---GA 546

Db 519 LTTSSGLTTHGGVTGKLFVADTYTGKLLKEIDTGTIPMAAPMTVEYGVQYIAVWAGWG 578

Qy 547 GGLTYGTQUNA 557

Db 579 GGYPFVPRISA 589

## RESULT 14

US-10-369-493-4929

/ Sequence 4929, Application US/10369493

/ Publication No. US20030233675A1

/ GENERAL INFORMATION:

/ APPLICANT: Cao, Yongwei

/ APPLICANT: Hinkle, Gregory J.

/ APPLICANT: Slater, Steven C.

/ APPLICANT: Goldman, Barry S.

/ APPLICANT: Chen, Xianfeng

/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

/ FILE REFERENCE: 38-10(52052)B

/ CURRENT APPLICATION NUMBER: US/10/369,493

/ PRIOR FILING DATE: 2003-02-28

/ PRIOR APPLICATION NUMBER: US 60/360,039

/ PRIOR FILING DATE: 2002-02-21

/ NUMBER OF SEQ ID NOS: 47374

/ SEQ ID NO 4929

/ LENGTH: 683

/ TYPE: PRT

/ ORGANISM: Burkholderia fungorum

US-10-369-493-4929

Query Match 16.7%; Score 513.5; DB 14; Length 683;

Best Local Similarity 26.9%; Pred. No. 2,4e-36;

Matches 158; Conservative 94; Mismatches 245; Indels 91; Gaps 20;

Qy 19 PAAFA-----CVPITDEL--ANPPAGEWINGRNOENYHSPLOI 59

Db 9 PAAFAAEVONSSSTTSSPVSTLQPVSGQDRAATSDMLNSGSAVQATYYGAQI 68

Qy 60 TADNVGLO--LWARGMEAGAVQVTMINDGVNLANPQDVQALDAQGLIEMHRQ 117

Db 69 NKTIVSKLRPAFIQTAVNE-SMETAVISNGVFIITTSFNHYAVADAATGKEFMHYK 127

Qy 118 LPAVATINAGQDRKRGVALYGTSLYFSSWMDNHLALDMETGOVVFVERSGEGDGLTSNT 177

Db 128 MGAITTT--CCGPNRGRGAIAGDRLVMGTLDKALVADAKGSLVMQTLADPREGT-SST 185

Qy 178 TGPVANGVIVAGSTCYSPYGC--FISGHSATGEELWNRHPIPOGEEGETW----- 230

Db 186 MAPTVVDGKVLIGT--NGEYIGRGLKAPDANGCOLMWFYTLPTGQEG--VMATKDA 241

Qy 231 -----GNDPEARMWTGVWGQITTDPTNLVFGSTGVGPASETGRTGG 271

Db 242 TGRDGRKIDAEKKQALAEKGGDFYKTLGGGVWMAPIADROTHVTFV---VWGNPSDLYG 298

Qy 272 T--PGTLYGTNTRFAVRPDTGEIWMRHQTLPRNNMQECTPEMMVANDVQSAEMEG 329

Db 299 AIRPQDNILY-TDLSVALDIDLGKTKMHQYFPHDWDLDVSPMLLDV----- 346

Qy 330 RAIPNNAATGERRVLTGAPCKTGTMSSPDAASGEFLWARDNTYNNIASIDEFTGLVTNE 389

Db 347 RDNNGRMIPG---VWHGQ--KTGHVYVHDBRATGRLI-----RVSQ--AMIPQENMTLPT 394

Qy 390 DAVLKELDEYDVCEPFLGSDWSSALNPTGIYFLPNNACDIYMDVQGEFSLDVYN 449

Db 395 AAGAML-----PGANGVWSSPMKPPDHTIRLVNANHOPPTTYVEDAAPGSGKLW 447

Qy 450 TSAATKLAPEGNNGRIDAIDISTGRTLSAERPAANYSPVLTAGVFNNGGTDYFR 509

Db 448 LGSAFKTILASQOMKLSAVVVDITGVAMDPKTEQPLIGVLTTAGGLVFNGBGGLFRA 507

Qy 510 LSQETGTLWQARLATVATGQALSYELDGYQYIAIGAGLTYGTQUNA 557

Db 508 FDSATGKKLWEPQCGAGVNAFVWVWKGQYIAVAAAG-QYARLQA 554

## RESULT 15

US-10-369-493-15969

/ Sequence 15969, Application US/10369493

/ Publication No. US20030233675A1

/ GENERAL INFORMATION:

/ APPLICANT: Cao, Yongwei

/ APPLICANT: Hinkle, Gregory J.

/ APPLICANT: Slater, Steven C.

/ APPLICANT: Goldman, Barry S.

/ APPLICANT: Chen, Xianfeng

/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

/ FILE REFERENCE: 38-10(52052)B

/ CURRENT APPLICATION NUMBER: US/10/369,493

/ PRIOR FILING DATE: 2003-02-28

/ PRIOR APPLICATION NUMBER: US 60/360,039

/ PRIOR FILING DATE: 2002-02-21

/ NUMBER OF SEQ ID NOS: 47374

/ SEQ ID NO 15969

/ LENGTH: 592

/ TYPE: PRT

/ ORGANISM: Xanthomonas campestris

US-10-369-493-15969

Query Match 15.7%; Score 482.5; DB 14; Length 592;

Best Local Similarity 26.4%; Pred. No. 1.1e-33;

Matches 165; Conservative 105; Mismatches 227; Indels 129; Gaps 28;

Qy 34 ANPPAGEWINGRNOENYHSPLOITADNVGQQLVW-----ARGMEAGAVQVTMHI 87

Db 14 SNP--DNWGVGRDPAFLTRHSPFLAENRDVYKULKMENKMTDTRHGG-----QGLVI 66

Qy 88 DGVNVL--ANDGVYQALDAQT---GDLIEMHRQLP---AVA---TNAQGRDRRGV 134

Db 67 GSIMYMAVAPNNV-VALDILASQDDGGVLMKYTPQODERSVAVAACCDTVN-----RGA 119

Qy 135 ALVGTSLYFSSWMDNHLALDMETGOVVFVERSGEGDGLTSNTGPIVANGVIAG--STC 193

Db 120 SYADGKLVFGLSGDVILADAKTAEVWQKLGHPDKEIT-TMAPIIADGKIYAGISGN 178

Qy 194 QYSPYGFISGHSATGEELWR-----NHPIPOGEEG--ETWGNDF 234

Db 179 EFGVLG-RVAAYNLADGQAWSCDAAGDKSICIGADFNKNAPOHGQGLDGLGVTFPND- 236

Qy 235 EARMWTG---VWGQITTDPTNLVFGSTGVGPASETGRTGGTLT-----GT 280



Db 237 --EMKRGGGAAMGWYSYDPKILLYGTGNPGLMSPSYR--CGKTSHEECNNGEHDNKM 292  
 QY 281 NTRFAVRPDGGEIYWRHQTLPNDWDOECTFEMWVANVDVQPSAEMEGLAIPNATGE 340  
 Db 293 MTLFARKIDTGEAVWGQKTPFDQMDYDGINETLVDLT-----DGKEV--PSVVQFD 344  
 QY 341 RRVLTGAPCKTGTWSPDASGEFLMARDNTNMTASID-ETGL-VTVNEDAVLKELDV 398  
 Db 345 R-----NGFAYVIDRRDGTLLRAHKFVPANWAEIRIDMKTGRPYKVAHSPL-ERGX 394  
 QY 399 EYDVCPTFLGPRDSSAALNP-DTGIYFLPLNACIDIMAVDQFSAIDVYNTSATAKLA 457  
 Db 395 KVQAFPSAMGGKQDQPCSVDPANSVAFPCGTNNHMELEPQERGNTWMLPYVFANVMK 454  
 QY 458 PGFEN-MGRIDALDIDSTGRTLWSAERPAANYSPVLTAGGVFNFGTDRYFRALSOETGE 516  
 Db 455 PNERGALGIYKAPFVWEGSKWEIKEKFPVWSGTLVTDGLVFYGTLDGWFRVVDKDTGK 514  
 QY 517 TLWGRLATVATGGAISYEIDGVOYIAI-----GAGGLTY 551  
 Db 515 KLWEMKLPISGIGNPPIAYKANHQYAVFSGIGWIGLPVAAGIDPADPYGALGAAGLAF 574  
 QY 552 GTQINAPLEAIDSTSVGNAIYVFAI 577  
 Db 575 GA-----GFDKIPLGWVHTFRI 592

Search completed: November 18, 2004, 05:23:37  
 Job time : 147 secs

***This Page Blank (usp10)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 18, 2004, 04:29:32 ; Search time 197 Seconds  
(without alignments)  
1691.076 Million cell updates/sec

Title: US-10-802-682-8  
Predict score: 3069  
Sequence: 1 MNPTTLRTSAVALTLTAPA.....AAHIDSTSVGNALTYVAPLQ 579

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 02.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1673	54.5	608	2 Q93RE9	Q93RE9 pseudogluc
2	680.5	22.2	575	2 Q96GX5	Q96GX5 bradyrhizob
3	629	20.5	554	2 Q89GY8	Q89GY8 bradyrhizob
4	624.5	20.3	557	2 Q6SEZ4	Q6SEZ4 uncultured
5	624.5	20.3	557	2 AAR38428	AAR38428 uncultured
6	624.5	20.3	724	2 Q6N4Z5	Q6N4Z5 rhodospseudo
7	624.5	20.3	724	2 CAE28629	CAE28629 rhodospseu
8	623	20.3	698	2 Q9KH03	Q9KH03 alcaligenes
9	612.5	20.0	690	2 Q8GR64	Q8GR64 pseudomonas
10	597	19.5	708	2 Q8ED COMTE	Q8ED COMTE pseudomonas
11	594	19.4	595	2 Q88JH0	Q88JH0 pseudomonas
12	594	19.4	695	2 Q9F9I2	Q9F9I2 pseudomonas
13	593.5	19.3	602	2 Q89GY2	Q89GY2 bradyrhizob
14	582	19.0	588	2 Q6W1W7	Q6W1W7 rhizobium s
15	582	19.0	588	2 AAQ87251	AAQ87251 rhizobium
16	574.5	18.7	757	1 DHET GLUTOX	DHET GLUTOX gluconobact
17	568	18.5	601	2 Q92WY9	Q92WY9 rhizobium m
18	565	18.4	629	2 Q9A048	Q9A048 methylobact
19	558	18.2	601	2 Q9EYV8	Q9EYV8 rhizobium m
20	556	18.1	691	2 Q9AP95	Q9AP95 pseudomonas
21	553.5	18.0	586	2 Q89D06	Q89D06 bradyrhizob
22	553	18.0	739	1 DHET ACEBU	DHET ACEBU pseudomonas
23	552	18.0	623	1 EXAA_PSEAE	EXAA_PSEAE pseudomonas
24	551.5	18.0	738	1 DHET ACEPO	DHET ACEPO pseudomonas
25	548.5	17.9	626	1 DHM1_METOR	DHM1_METOR methylobact
26	543	17.7	582	2 Q89YX1	Q89YX1 bradyrhizob
27	542.5	17.7	622	2 Q75Q65	Q75Q65 pseudomonas
28	542.5	17.7	622	2 BAD1158	BAD1158 pseudomonas
29	542	17.7	750	2 Q6SFC4	Q6SFC4 uncultured
30	542	17.7	750	2 AAR38298	AAR38298 uncultured
31	541.5	17.6	631	2 Q88JH5	Q88JH5 pseudomonas

32	538	17.5	601	2 P71509	P71509 methylobact
33	537	17.5	623	2 Q9AGW3	Q9AGW3 pseudomonas
34	535.5	17.4	626	1 DHM1_METEX	DHM1_METEX methylobact
35	533.5	17.4	631	2 Q8VPS5	Q8VPS5 pseudomonas
36	530	17.3	750	2 Q8RTS8	Q8RTS8 uncultured
37	511.5	16.7	742	2 Q53362	Q53362 acetobacter
38	504.5	16.4	633	2 Q24759	Q24759 hyphomicrob
39	498	16.2	631	1 DHM1_PARDE	DHM1_PARDE paracoccus
40	493.5	16.1	661	2 Q8PPN7	Q8PPN7 xanthomonas
41	491.5	16.0	742	1 DHET ACEAC	DHET ACEAC acetobacter
42	491.5	16.0	742	2 Q8KZS8	Q8KZS8 acetobacter
43	488.5	15.9	600	1 XOXF_PARDE	XOXF_PARDE paracoccus
44	471	15.3	695	2 Q934G0	Q934G0 pseudomonas
45	468.5	15.3	599	2 Q8RMH5	Q8RMH5 methylobact

ALIGNMENTS

RESULT 1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

```

Db      368 TRKVLGVPECKTAVANQFPAKTDGYFWSKATVEQNSIASIDDTGLVTVEMDILKEPGKT 427
Qy      400 YDQCPFFLGGSDSSALNPDGTGYFLPLNNAACYDMAYDOERSALDVNTSATAIAGB 459
Db      428 YNCCPFFLGGSDMPGAGLPEKSNLYIPLSNACYDMARTETKTPADVNTDAILVLANG 487
Qy      460 FENMERIDAIDISTGRTLWSAERPAAVNSPVLSTAGGVFNQGTDRYPALSOETGETILM 519
Db      488 KTMNGRVDAIDLATGETKWSYETRAALYDPVLTGTGDLVFEVGGIDRDPALDAESGKEVW 547
Qy      520 QARLATVATGQATSYELDGVQYIAIGAGGLTYGTQANAPLAIBDSTVGNAIYVPAALPQ 579
Db      548 STRLPAGVSGITTSYSDGNQYVAVVSGG-SLGGPFGPTTPVDSDASANGIYVPALE 606

```

## RESULT 2

```

Q89GX5 PRELIMINARY; PRT; 575 AA.
ID 089GX5
AC 089GX5
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, last annotation update)
DE B16220 protein.
GN OrderedLocustNames=b16220;
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idegawa K., Iriyuchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsurutoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005957; BAC51485.1; -.
DR HSSP; Q924J7; 1FUG.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002372; PQQ_repeat.
DR InterPro; IPR011047; Quin_alc_DH_like.
DR Pfam; PF01011; PQQ; 3.
DR SMART; SM00564; PQQ; 6.
KW Complete proteome.
SQ SEQUENCE 575 AA; 63421 MW; F428035287AD54D CRC64;

```

Query Match 22.2%; Score 680.5; DB 2; Length 575;  
 Best Local Similarity 32.6%; Pred. No. 2.5e-38;  
 Matches 197; Conservative 99; Mismatches 226; Indels 83; Gaps 27;

```

Qy      12 AVLLLTAPAF-----QVPIITDELLANPAGEMVINGNOEYRHSPLTQITADNVGQL 67
Db      13 AFTCLASTAAGAGRIENYSPTAQRLENPEPSNMWMLRRTYDQGSPLDQITNSVKG 72
Qy      68 QLVW--ARGWEAGAVQVTPIIHIDGVVYLANPGDVIQALDAQTGLIWEHRRLPAVATLN 125
Db      73 TPVWTFATGVVEGH-EAPPIVNGVMFVATPMGVQVALANAKTDEYMRKRGQP--DDL 129
Qy      126 AQGDRKRGVALYGTSLYSFSSWDNLIALDMETGOVVFVDERGSGEGLTSNTGPIVANG 185
Db      130 QLPHTSGVGLMDEKYLATTDDHVVALDAKTGVWMDTKVQDKKG-QYMTLMLPLIVG 188
Qy      186 -VIVAGSTQVSPGFCISGDSATGELWRNHFIPQGEEDETW-GNDPE---ARWM 239
Db      189 KVIIVGSGGEGFGRG-VVAIVDAKDGKELMRTYTIIGEGEPGHDTMGSDMKNGGSSAM 247
Qy      240 TGVWQITVDPVNNLVFVSGTGVGP-ASEFQRTGGTLYGINTRAVAPDPTGEIYWRQ 298
Db      248 TG-----NYDKDKTXYIYVGVGNAAFWPGEETH---PGDNLY-TSSVLALDPNNKIKTYHQ 298

```

```

Qy      299 TLPRDNDQECTEFMYAVNDVQPSAEMEGRLAINPAAIGERR---VLTGAPCKTGM- 354
Db      299 YHONDSMDWEVEAEAPML--IDLQRDQ-----RSIKSLVHGDRDAIFWVLERTPTKINYVA 351
Qy      355 -WSPDASGEFLMARDNTYNNMISID-ETGLVTVNEDAVLKELDVEYDVCPTFLGGRDW 412
Db      352 GMPFVS-----IDVWKGIDAETGKPIV-DPAKPIVIGKREVCPSLMGSKDM 397
Qy      413 SSAALNPDGTGYFLPLN-NAC---YDIMAVDQERSALDVYNTSATKALPGEENGRID 467
Db      398 PSAVYSQKTGLIVYVPAENECGCGFTGKVKALKGELMLGTRPEDIGLKTGPDHFELO 457
Qy      468 AIDISTGRTLWSAERPAAV--YSPVLTSTAGGVFNQGT-DRYPALSOETGETIMOARLAT 525
Db      458 AMDPVYTGKWKQHNFPKSQLFGSYTATAGDLIFAGGINDNPPAPNAKTELLMECKTNS 517
Qy      526 VATGOAISYELDGVQYIAIGAG-GL-----TYGTQANAPLAIBDSTVGNAI 572
Db      518 GIMGMFVSYEIDGTQYIAIQSGWGVDAQRIDQALVTNNIGIEANVP-----QGGVI 568
Qy      573 YVPA 577
Db      569 WVPAL 573

```

## RESULT 3

```

Q89GY8 PRELIMINARY; PRT; 554 AA.
ID 089GY8
AC 089GY8
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, last annotation update)
DE Exa protein.
GN Name=exa; OrderedLocustNames=b16207;
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idegawa K., Iriyuchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsurutoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005957; BAC51472.1; -.
DR HSSP; Q924J7; 1FUG.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002372; PQQ_repeat.
DR InterPro; IPR011047; Quin_alc_DH_like.
DR Pfam; PF01011; PQQ; 2.
DR SMART; SM00564; PQQ; 6.
KW Complete proteome.
SQ SEQUENCE 554 AA; 60554 MW; CB8B797BBAD9B061 CRC64;

```

Query Match 20.5%; Score 629; DB 2; Length 554;  
 Best Local Similarity 30.5%; Pred. No. 8.4e-35;  
 Matches 184; Conservative 101; Mismatches 229; Indels 90; Gaps 23;

```

Qy      10 SAALLLTAPAFQVPTITDELLAN-PPAGEMVINGRNOENRHSPLTQITADNVGQL 68
Db      7 AAILVMASTVANNQT--TEQLVKGATDTSNVNTVMGYNLRFTSLNINKDTVNNLV 63
Qy      69 LVWARGM-EAGAVQVTPIIHIDGVVYLANPGDVIQALDAQTGLIWEHRRLPA-VATLNA 126
Db      64 PWNVYSPRDBSBSQPLVYGVVYYS-RNATAVDAKXGKQWKSXIYPAETPRIVC 122
Qy      127 QGDRKRGVALYGTSLYSFSSWDNLIALDMETGOVVFVDERGSGEGLTSNTGPIVANG 186

```

```
Db 123 CGIIRRGALYDGVFFETITDANVIALDADGKELWQKADIKKEG-SMTTAPVADGV 181
Qy 187 IVAG-STCQSPYGCIFSGHDSATGEELWNNHPIPOGEEGDETWGNDP-----EARMNT 240
Db 182 VIRGISCAEFGTRG-FIDGWDPAFGKILMRTHSPSPDEFGDWTGDKTGKGGSGSTWIT 240
Qy 241 GWVGQITVDVTNLVFGSTGVGPASBETGPGGTLYGTNTRFAVRPDITGEIWMHQTL 300
Db 241 G-----SYDELNTVTVGIGNPGPFNSAVR-PQDNLY-TCSVLADPRTGKIKMHYQPS 292
Qy 301 PRDNMOECTFEMWVA--NVDPQSAEMEGRAINPNAATGERRVLTGAPCKTGTWMSFD 358
Db 293 PNPFPDYDVAEVLADMNTECKPT-----KVLMDAN-RNGFFVYLD 333
Qy 359 AASGEFLMADTNTYNNIASID-ETGLVTVNEDAVLKEIDVEYD-----YCFPLGG 409
Db 334 RTNGKLLAANPYKVNATGVDMTKGRPI-----ETDVSXDAEKKVTVYPSILGG 385
Qy 410 RDWSSALNPDGTGYFLPLNNACYDIWVDOEFSALDVNTSATKLAPGFENM----- 463
Db 386 KNMEPMSFNPQTGL-----AVANTLAFGKXKAEVY-----TFQGEWYLGMDLTDPM 433
Qy 464 -----GRIDAIDISTGRTLMSAERPAANYSPVLSTAGGVVNGTDRYFRALSOETGB 516
Db 434 EFGDGAHGLKALDPMTGKAKWEAPADIPRFGSLSTAGGVVFTGALTGFEAFDADTGK 493
Qy 517 TLWQARLATVATGQAISELYEDGVYTAIGAGLYTGQNLAPLA-BAIDSTVSNAIYF 575
Db 494 KLMQFGQSGSIEQGPVTWQDDVQYIAVYSG---YGGVYISLFGDERLAKVPPGSLWVF 550
Qy 576 ALPQ 579
Db 551 AVKQ 554
```

## RESULT 4

```
ID Q6SE24 PRELIMINARY; PRT; 557 AA.
AC Q6SE24;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE PQQ enzyme repeat domain protein.
GN ORFNames=EBAC080-L028H02.93;
OS uncultured bacterium 582.
OC Bacteria; environmental samples.
OX NCBI_Taxid=257402;
RN [1]
RP SEQUENCE FROM N.A.
RA Delong E.F.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Delong E.F.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL: AY458649; AAR38428.1; -.
DR InterPro: IPR002372; PQQ_repeat.
DR InterPro: IPR011047; QunA_alc_DH_like.
DR Pfam: PF01011; PQQ; 2.
DR SMART: SM00564; PQQ; 7.
SQ SEQUENCE 557 AA; 61170 MW; 10A36DDEF2D15A08 CRC64;
```

Query Match 20.3%; Score 624.5; DB 2; Length 557;

Best Local Similarity 32.3%; Pred. No. 1,7e-34; Matches 189; Conservative 82; Mismatches 250; Indels 65; Gaps 22;

```
Qy 12 AVLLITAPAAFAQVPTITDELANPAGEWINGRQENYRHSPLQTITADNVGQLQVW 71
Db 15 ALSLIAPETTLAOST--SDLINDTKDTTLLTYGMYGGQTRHSGLAGINDLNIGRLRPEW 72
Qy 72 ARGMEAGAVVT-PMIHGQVYLANPGDVYQALDAQTGLIIV---EHRRLPVAATLNA 126
```

```
Db 73 TYSLSDTRGQETFFLIHDKQVYVTTASTM-ALDLVTRGQIKTKLVDPADTPRVACC-- 129
Qy 127 QGDRRGVALYGTSLYFSSWMDNHLIALDMETQGVVDFVERSGEDGLTSNTTGPVANGV 186
Db 130 -GIVRGAVALYEGKVFRTLLDAHVIALDAETGNELMRQSDIKHTGY-SMTVAPWMAAGV 187
Qy 187 IVAG-STCQSPYGCIFSGHDSATGEELWNNHPIPOGEEGDETWGNDP-----EARMNT 240
Db 182 VIRGISCAEFGTRG-FIDGWDPAFGKILMRTHSPSPDEFGDWTGDKTGKGGSGSTWIT 240
Qy 241 GWVGQITVDVTNLVFGSTGVGPASBETGPGGTLYGTNTRFAVRPDITGEIWMHQTL 300
Db 241 G-----SYDELNTVTVGIGNPGPFNSAVR-PQDNLY-TCSVLADPRTGKIKMHYQPS 292
Qy 301 PRDNMOECTFEMWVA--NVDPQSAEMEGRAINPNAATGERRVLTGAPCKTGTWMSFD 358
Db 293 PNPFPDYDVAEVLADMNTECKPT-----KVLMDAN-RNGFFVYLD 333
Qy 359 AASGEFLMADTNTYNNIASID-ETGLVTVNEDAVLKEIDVEYD-----YCFPLGG 409
Db 334 RTNGKLLAANPYKVNATGVDMTKGRPI-----ETDVSXDAEKKVTVYPSILGG 385
Qy 410 RDWSSALNPDGTGYFLPLNNACYDIWVDOEFSALDVNTSATKLAPGFENM----- 463
Db 386 KNMEPMSFNPQTGL-----AVANTLAFGKXKAEVY-----TFQGEWYLGMDLTDPM 433
Qy 464 -----GRIDAIDISTGRTLMSAERPAANYSPVLSTAGGVVNGTDRYFRALSOETGB 516
Db 434 EFGDGAHGLKALDPMTGKAKWEAPADIPRFGSLSTAGGVVFTGALTGFEAFDADTGK 493
Qy 517 TLWQARLATVATGQAISELYEDGVYTAIGAGLYTGQNLAPLA-BAIDSTVSNAIYF 575
Db 494 KLMQFGQSGSIEQGPVTWQDDVQYIAVYSG---YGGVYISLFGDERLAKVPPGSLWVF 550
Qy 576 ALPQ 579
Db 551 AVKQ 554
```

## RESULT 5

```
ID AAR38428 PRELIMINARY; PRT; 557 AA.
AC AAR38428;
DT 02-MAR-2004 (Tremblrel. 27, Created)
DT 02-MAR-2004 (Tremblrel. 27, Last sequence update)
DT 02-MAR-2004 (Tremblrel. 27, Last annotation update)
DE PQQ enzyme repeat domain protein.
GN EBAC080-L028H02.93.
OS uncultured bacterium 582.
OC Bacteria; environmental samples.
OX NCBI_Taxid=257402;
RN [1]
RP SEQUENCE FROM N.A.
RA Delong E.F.;
RL "Monterey Bay Coastal Ocean Microbial Observatory environmental clone sequencing.";
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Delong E.F.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL: AY458649; AAR38428.1; -.
DR InterPro: IPR002372; PQQ_repeat.
DR InterPro: IPR011047; QunA_alc_DH_like.
DR Pfam: PF01011; PQQ; 2.
DR SMART: SM00564; PQQ; 7.
SQ SEQUENCE 557 AA; 61170 MW; 10A36DDEF2D15A08 CRC64;
```

Query Match 20.3%; Score 624.5; DB 2; Length 557;

Best Local Similarity 32.3%; Pred. No. 1,7e-34; Matches 189; Conservative 82; Mismatches 250; Indels 65; Gaps 22;

```
Qy 12 AVLLITAPAAFAQVPTITDELANPAGEWINGRQENYRHSPLQTITADNVGQLQVW 71
Db 15 ALSLIAPETTLAOST--SDLINDTKDTTLLTYGMYGGQTRHSGLAGINDLNIGRLRPEW 72
Qy 72 ARGMEAGAVVT-PMIHGQVYLANPGDVYQALDAQTGLIIV---EHRRLPVAATLNA 126
Db 73 TYSLSDTRGQETFFLIHDKQVYVTTASTM-ALDLVTRGQIKTKLVDPADTPRVACC-- 129
Qy 127 QGDRRGVALYGTSLYFSSWMDNHLIALDMETQGVVDFVERSGEDGLTSNTTGPVANGV 186
```

```

Db      130  |G|V|N|G|V|A|L|Y|E|K|V|E|R|T|L|D|A|V|I|A|L|D|E|T|G|E|L|M|R|T|G|S|I|D|H|K|Y|G|S|N|T|A|P|A|M|A|G|V| 187
      187  |V|A|G|S|T|C|O|Y|S|P|Y|G|C|T|S|G|H|S|A|T|G|E|L|P|R|N|H|P|O|P|G|E|G|D|E|T|W|G|N|D|E|A|R|-----W| 238
      188  |L|I|G|I|S|G|E|G|A|I|G|G|V|L|E|G|Y|D|P|D|T|G|V|R|V|W|R|Y|T|I|P|A|P|G|E|G|S|E|T|W|D|G|D|A|W|R|G|A|P|T| 246
      239  |M|T|G|W|G|I|T|D|P|T|N|I|V|F|G|S|T|G|V|P|A|S|E|T|G|T|G|T|Y|G|T|N|T|R|A|V|A|P|D|G|E|I|V|W|R|Q| 298
      247  |L|I|D|-----S|Y|D|E|L|N|V|F|W|G|T|G|N|S|S|N|A|A|T|R|E|D|N|I|Y|A|S|I|L|A|D|P|K|G|T|I|K|H|N|Q| 298
      299  |T|L|R|D|N|W|D|E|C|T|E|N|N|V|A|N|D|V|O|P|S|A|E|G|L|A|I|N|E|N|A|A|T|G|R|R|V|T|G|A|P|C|K|T|G|T|W|S|D| 358
      299  |T|T|P|N|D|E|F|D|T|V|E|L|H|A|D|I|D|-----G|R|K|V|L|Q|A|N|G|N|G|E|F|Y|V|A|D| 337
      359  |A|A|S|G|E|L|A|R|D|N|T|N|N|M|A|S|I|D|E|T|G|I|V|N|E|D|A|V|I|K|E|L|D|V|E|D|V|C|P|F|L|G|R|W|S|S|A| 416
      338  |R|A|G|E|L|I|A|A|K|F|V|D|K|I|N|A|D|S|I|D|E|T|R|P|R|E|T|D|V|A|K|A|S|G|E|I|T|T|W|P|A|L|G|K|W|S|P|A| 397
      417  |L|N|D|T|G|I|V|E|L|P|L|N|N|A|C|Y|D|I|A|V|D|E|F|S|A|L|D|V|N|T|S|A|T|A|K|A|P|E|N|N|G|R|I|D|A|I|D|I|S|T|G|R|T| 476
      398  |Y|N|K|T|T|I|P|A|N|T|K|F|G|H|Y|K|A|V|E|P|Q|Y|R|P|G|V|F|F|G|A|E|F|S|W|D|P|D|G|R|G|E|L|R|A|F|D|P|T|G|V| 456
      477  |L|N|S|A|R|P|A|N|Y|S|P|V|S|T|A|G|V|V|N|G|T|D|R|Y|F|R|A|L|S|E|T|G|E|T|I|L|Q|R|A|L|A|V|A|T|G|A|I|S|Y|E|L| 536
      457  |K|W|S|D|S|I|A|P|R|Y|S|G|V|S|T|G|G|L|V|F|T|G|N|W|D|E|F|E|A|F|N|A|E|T|D|K|W|T|Y|S|T|G|S|G|I|I|G|P|V|T|W|R| 516
      537  |D|G|V|Y|A|I|-----G|A|G|G|-----L|T|Y|G|T|Q|L|N|A|P|L|A|E|I|D|S|T|S|V|G|N|A|I|V|F|A|L| 577
      517  |E|G|Q|Y|I|T|V|A|N|G|S|G|V|S|L|F|S|G|D|-----E|R|L|A|S|V|A|G|N|I|W|T|F|S|I| 555

RESULT 6
ID 06N425 PRELIMINARY; PRT; 724 AA.
AC 06N425;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE Alcoh1 dehydrogenase precursor (EC 1.-.-.-).
GN OrderedLocustNames=RPJA188;
OS Rhodospseudomonas palustris.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Rhodospseudomonas.
OX NCBI_TaxID=1076;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CGA009 / ATCC BAA-98;
RX PubMed=14704707; DOI=10.1038/nbt923;
RA Larimer F.W., Chain P., Hauser L., Lamerdin J.E., Malfatti S., Do L.,
RA Land M.L., Pelletier D.A., Beatty J.T., Lang A.S., Tabita F.R.,
RA Gibson J.L., Hanson T.E., Bobst C., Torres y Torres J.L., Peres C.,
RA Harrison F.H., Gibson J., Harwood C.S.;
RT "Complete genome sequence of the metabolically versatile
RT photosynthetic bacterium Rhodospseudomonas palustris."
RL Nat. Biotechnol. 22:55-61(2004).
EMBL: BX572603; CAE28629.1; -.
CG: GO:0016491; F:oxidoreductase activity; IBA.
DR InterPro: IPR00345; CytC_heme_BS.
DR InterPro: IPR009056; Cytochrome_C.
DR InterPro: IPR003088; Cyt_C1.
DR InterPro: IPR002372; PQQ_repeat.
DR InterPro: IPR011047; Quin_alc_DH_like.
DR Pfam: PF00034; Cytochrom_C_1.
DR SMART: SM00564; PQQ_5.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
KM Complete proteome; Oxidoreductase; Signal.
FT SIGNAL 1 45 Potential.
RL Nat. Biotechnol. 22:55-61(2004).
SQ SEQUENCE 724 AA; 78144 MW; 474551BA6FCBCE98 CRC64;
Query March 20.3%; Score 624.5; DB 2; Length 724;
Best Local Similarity 29.5%; Pred. No. 2; Se-34;

```

```

Matches 184; Conservative 100; Mismatches 249; Indels 91; Gaps 24;
      8  |R|T|S|A|V|I|-----L|I|P|A|R|A|P|A|O|-----V|T|P|I|D|E|-----L|L|A|N|-----D|P|A|G|E|W|I|Y|G|R|N|G|E|N| 50
      26  |R|L|G|A|A|L|A|A|L|I|P|A|V|A|A|A|D|S|K|S|A|H|I|A|V|G|A|V|D|S|A|I|V|A|N|V|T|K|D|M|P|E|Y|G|D|Y|A|E| 85
      51  |Y|R|H|S|P|L|Q|I|T|A|D|N|V|G|Q|L|V|W|-----A|R|G|E|A|G|A|V|O|T|P|H|I|D|G|V|N|Y|L|A|N|P|D|V|I|O|A|L|D|A|Q|T| 108
      86  |T|R|F|S|K|D|Q|I|E|N|V|K|Q|L|G|L|O|N|S|Y|S|L|S|E|R|G|-----V|E|A|T|P|V|V|D|G|I|V|V|A|S|V|V|H|A|I|D|R|T|G| 144
      109  |D|L|I|E|H|R|Q|L|P|A|V|A|T|L|N|A|O|D|R|-----K|R|G|V|A|L|Y|G|S|L|F|S|S|W|H|L|A|L|D|E|T|G|Q|V|R| 162
      145  |K|K|L|T|F|D|P|K|V|D|-----H|S|K|Y|R|G|C|C|D|V|N|G|V|A|L|Y|K|K|Y|V|G|A|Y|D|R|L|I|L|D|A|A|T|S|K|A| 199
      163  |D|V|E|R|S|G|E|D|G|I|S|N|T|T|P|I|V|A|N|G|I|V|A|G|S|-----T|C|O|Y|S|P|Y|G|C|T|S|G|H|S|A|T|G|E|L|P|R|N|H|P| 220
      200  |E|I|D|L|I|D|H|E|S|Y|I|T|G|A|P|R|V|E|N|G|V|V|I|N|G|C|A|E|Y|G|A|R|G|-----V|Y|A|V|A|D|E|T|G|Q|A|R|W|R|F|V|P|G| 258
      221  |-----Q|G|E|D|E|F|W|G|N|D|F|E|R|K|W|T|G|-----V|W|Q|I|T|D|P|T|N|I|V|F|G|S|T|G|V|P|A|S|E|T| 268
      259  |D|P|S|K|P|E|D|S|E|K|A|K|T|W|-----D|P|A|G|K|W|L|N|G|G|G|T|A|D|T|I|T|D|P|L|N|L|I|V|G|T|G|N|G|S|P|W|R|H| 317
      269  |Q|R|T|P|G|G|I|V|G|T|R|F|A|R|P|D|T|G|I|V|W|R|H|Q|L|P|R|D|N|W|D|E|C|T|E|N|N|V|A|N|-----V|D|V|O|P|S|A|E|W| 326
      318  |L|R|S|P|A|G|D|N|L|Y|A|S|I|V|A|L|N|A|D|T|G|Y|V|W|H|Y|G|E|T|G|D|W|D|Y|S|T|Q|P|H|I|A|I|D|G|K|P|----- 373
      327  |E|G|R|A|I|N|P|N|A|T|G|R|R|V|L|T|G|A|P|C|K|T|G|T|W|S|F|D|A|S|E|F|L|A|R|D|N|T|N|T|A|S|I|D|E|T|G|L|V| 385
      374  |-----R|K|Y|I|L|A|P|-----K|N|G|F|F|V|I|D|T|D|G|F|I|S|A|K|F|V|D|V|N|A|T|G|Y|D|S|N|G|R|P|I| 418
      386  |T|V|N|E|D|A|V|L|K|E|L|D|V|E|D|V|C|P|F|L|G|R|D|W|S|A|L|N|P|D|T|G|I|V|E|L|P|L|N|N|A|C|Y|D|I|-----M|A|V|D|Q|----- 440
      419  |E|V|P|E|-----A|R|S|A|D|S|F|D|I|P|G|Y|G|A|H|W|H|P|S|F|P|Q|I|G|V|L|P|A|Q|V|P|V|N|I|T|G|S|K|A|L|Q|N|K| 475
      441  |-----E|F|S|A|L|D|V|N|T|S|A|T|A|K|A|P|E|N|N|-----G|R|I|D|A|I|D|I|S|T|G|R|T|L|S|A|R|P|A|N|Y|S|P|V|L|S|T|R|A| 494
      476  |M|E|R|F|S|T|G|T|G|N|V|G|F|L|N|A|V|P|-----K|N|L|P|F|R|L|A|V|A|P|V|O|Q|E|W|R|A|E|Y|S|P|W|N|G|I|L|T|T|A| 534
      495  |G|G|V|E|N|G|T|G|R|Y|P|A|L|S|E|T|G|E|T|I|L|Q|R|A|L|A|V|A|T|G|A|I|S|Y|E|L|D|G|V|Y|A|I|G|A|G|-----G|L|Y|G|T| 553
      535  |G|N|L|V|Q|G|T|A|D|R|F|V|A|N|A|K|G|E|K|I|W|S|P|L|G|T|G|A|V|A|P|A|T|T|W|D|Q|V|S|I|A|V|G|W|G|V|F|G|I| 594
      554  |Q|L|N|A|P|L|A|E|I|D|S|T|S|V|G|N|A|I|V|F|A|L| 577
      595  |S|A|R|A|T|E|R|E|A|-----P|G|V|Y|T|F|A|V| 612

RESULT 7
ID CAE28629 PRELIMINARY; PRT; 724 AA.
AC CAE28629;
DT 02-MAR-2004 (TReMBLrel. 27, Created)
DT 02-MAR-2004 (TReMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TReMBLrel. 27, Last annotation update)
DE Alcoh1 dehydrogenase precursor (EC 1.-.-.-).
GN RPA188.
OS Rhodospseudomonas palustris.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Rhodospseudomonas.
OX NCBI_TaxID=1076;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CGA009 / ATCC BAA-98;
RX PubMed=14704707;
RA Larimer F.W., Chain P., Hauser L., Lamerdin J., Malfatti S., Do L.,
RA Land M.L., Pelletier D.A., Beatty J.T., Lang A.S., Tabita F.R.,
RA Gibson J.L., Hanson T.E., Bobst C., Torres y Torres J.L., Peres C.,
RA Harrison F.H., Gibson J., Harwood C.S.;
RT "Complete genome sequence of the metabolically versatile
RT photosynthetic bacterium Rhodospseudomonas palustris."
RL Nat. Biotechnol. 22:55-61(2004).
EMBL: BX572603; CAE28629.1; -.
SQ Oxidoreductase; Signal.

```

FT SIGNAL 1 45 Potential.  
SQ SEQUENCE 724 AA; 78144 MW; 474551BA6FBC98 CRC64;  
Query Match 20.3%; Score 624.5; DB 2; Length 724;  
Best Local Similarity 29.5%; Pred. No. 2,5e-34;  
Matches 184; Conservative 100; Mismatches 249; Indels 91; Gaps 24;  
QY 8 RTSAAYLV---LTPAPAF---VTPITDE---LLAN--PPAGWINGYNGNEN 50  
DB 26 RLGAAALMALTLTPAALAAQDSKSAHITAVGADSAIVANVTTCMDSEYGDYAE 85  
QY 51 YRHSPLTQITADNVGQIQVW--ARQNEAGAVQVTPMHDGVVYLANPGDVIQALDAQTG 108  
DB 86 TRFSKIDQITENTENVKQIQMSYSLGSERG--VEATPVVDGIMYVIAVMSVVAIDTRIG 144  
QY 109 DLWEHRRLPAPATLNAQDR-----KRGVALYGTSLFSSWNHLLALDMEGQVVF 162  
DB 145 KXIMTDPKVD---HSGYRGCCDVVNGVALYKGVGVGADRLIALDATSCKAM 199  
QY 163 DVERGSGEDGLTNTGPIVANGVIYAGS--TCQSPYGCPISGHDSATGSELMRNHFIP- 220  
DB 200 EIDTLIDHEHSYITITGAPRVFNGKVITGNAGAEYARG--YVTAAYDAETGQAMRWFTVPG 258  
QY 221 -----QGEEDGDEWNGDFFARWMTG---VMQIITDPVTNLVPGSTGVGPASET 268  
DB 259 DPKPEDESEMEKAATW--DPAGKMWLNGGGGTAMDITTFDPDLNLIYGTGNGSEFWNH 317  
QY 269 QRGTPGGTLYGTNTPFAVRPDTEIIVMRHOTLPRDNWDOCEFEWVAV--VDVQPSAM 326  
DB 318 LRBPAGSDNLYLASYALNADTRKYVWHYETPGDMWDYTSQPMILADIDKRI----- 373  
QY 327 EGIKALNPATGERKVLTPGCKTGTMSFDAASGEFLMARDTNTNMILASIDETGL-V 385  
DB 374 -----RKYLHAP--KNGFFVIDRTDKFISAKNFVNVNATGDSNGRPI 418  
QY 386 TVNEDAVLKEIDVEYDVCFTFLGARDSSALNPDGIYFLPLNNAQVDI---MAVDO-- 440  
DB 419 EYVE---ARADSPFADIPGYPGAMHMHMSFNPQIGLYLPRQGVPPVNLGSEKALTOK 475  
QY 441 ---EESALDVNTSATKLAPEGENN--GRIDAIDISTGRTLSAERPANYSPYLSTA 494  
DB 476 MEFKFGSTGMVNGFTLNAPV--KNLPFGRLVAMPVQCKEVRABEYVSPMNGGTLTA 534  
QY 495 GGVVNGGDRYRRAISOETGELMQARLATATGALISYELDGVYIAGG--GLTYGT 553  
DB 535 GNLVFOGTADGRVAVNAKTEKTLMESPLGTGAVAAPATYWDGVQVSIAGWGVFGI 594  
QY 554 QLNAPLAEAIDSTVGNAIYFAL 577  
DB 595 SARATEEA-----PGTVYTFVAV 612  
RESULT 8  
Q9KH03 PRELIMINARY; PRT; 698 AA.  
ID Q9KH03  
AC Q9KH03; (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)  
DT 01-OCT-2000 (TREMblrel. 26, last annotation update)  
DT 01-MAR-2004 (TREMblrel. 26, last annotation update)  
DS Tetrahydrofurfuryl alcohol dehydrogenase.  
OS Alcaligenes eutrophus (Ralstonia eutropha).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Burkholderiaceae; Wauteria.  
OC NCBI\_TaxId=510;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21125557; PubMed=11222593;  
RA Zarni G., Schrader T., Andreesen J.R.;  
RT "Catalytic and molecular properties of the quinoxinoprotein  
tetrahydrofurfuryl alcohol dehydrogenase from Ralstonia eutropha  
strain Bo.";  
RT J. Bacteriol. 183:1954-1960(2001).  
RL J. Bacteriol. 183:1954-1960(2001).  
DR EMBL; AF277373; AAF86335.1; -.

DR HSP: Q46444; 1K30.  
DR GO; GO:0005746; C:mitchondrial electron transport chain; IEA.  
DR GO; GO:0005489; E:electron transporter activity; IEA.  
DR GO; GO:0006118; E:electron transport; IEA.  
DR InterPro; IPR000345; CyC\_heme BS.  
DR InterPro; IPR009056; Cytochrome\_c.  
DR InterPro; IPR003088; CyC\_C1.  
DR InterPro; IPR002372; PQQ\_repeat.  
DR InterPro; IPR01047; Quin\_alc\_DH\_like.  
DR Pfam; PF00034; Cytochrome\_c; 1.  
DR Pfam; PF01011; PQQ; 4.  
DR SMART; SM00564; PQQ; 5.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN 1.  
SQ SEQUENCE 698 AA; 75471 MW; 9C87F180BD12F82 CRC64;  
Query Match 20.3%; Score 623; DB 2; Length 698;  
Best Local Similarity 29.8%; Pred. No. 3e-34;  
Matches 182; Conservative 99; Mismatches 249; Indels 80; Gaps 24;  
QY 10 SAAVILTPAPAF-AQVPIPD--ELLANPPAG--EWINGRNOENYHSPLOITADNV 64  
DB 14 AAASVALPAPAFGANAARAVDGAIRAN--EAGTPNPSYGLDYAEIRFSLEQVNAAGNV 72  
QY 65 GOLQVWARGMA--GAVVTPMHDGVVYLANPGDVIQALDAQTGDLIWEHRRLPAPAT 123  
DB 73 RNLGLAMSYDLESIRGEATPLVVDGVVYSAFMSVVAIDARTGKRLMTYDPPVPRDQ 132  
QY 124 LNAQGD--KRGVALYGTSLFSSWNHLLALDMEGQVFPDVERGSGEDGLTNTGPIV 182  
DB 133 YKGCDDVYVRGVALYQGVFPGAFDGRVALDAITGKRWQDPIVDRSKSYITITGAPRV 192  
QY 183 ANG-VIYAGSTCQYSPYGCPISGHDSATGSELMRNHFIP---QGE-----EEDETWGN 232  
DB 193 YNGKYLINGAGAEYGVG--YITAYDAETGQKQWYVTPCDPAPFENEMAKAAATW--D 250  
QY 223 DFEARWMTG---VMQIITDPVTNLVPGSTGVGPASETGRRPG--TYGNTTFRAVR 287  
DB 251 PSKKTWINGGGGYVWNTAFPELNLMTIGTGNAPMSRKLRSKGGDNLTAASV--VALN 309  
QY 288 PDTGEIYMRHOTLPRDNWDOCEFEWVAV--VDVQPSAEMEGRLAIPNATGERVLT 345  
DB 310 PDTGVVWHYETPGDMWDYTSQDILADKIDGP-----RKYL 351  
QY 346 GAPCKTGTMSFDAASGEFLMARDTNTNMILASIDETG--LTVNEDAVLKEIDVEYDVC 403  
DB 352 HAP--KNGFFVIDRTNGKFIKAFVDVNAAGYDKRGPVETPOADTSGPA---DVV 406  
QY 404 PTFLGARDSSALNPDGIYFLPLNNAQVDIMAVDOFSALDVNTSATKL----- 456  
DB 407 GPGFAMHMHMSFHPKLGAFIFAGHYPL--TLADNKEW-----VHNKDSFEARHGVGN 461  
QY 457 -----APGFENGRIDAIDISTGRTLSAERPANYSPYLSTAGGVFNGGTDYFR 508  
DB 462 LGMVLNAPRPSKPMGRILAMPPLAKQAVMRHDAGPMNGGTLTAIYNLVFOGTADRLV 521  
QY 509 ALSQETGELMQARLATATGALISYELDGVYIAGG--GLTYGTQNLNPLAIAIDSTS 567  
DB 522 AYHATGKLMQALPTGSGVAAPVYTLIDRQIYSAVANGGVYG-----LSGRGSDRQ 575  
QY 568 VGNAIYFAL 577  
DB 576 AGRVYTFVL 585  
RESULT 9  
Q9GR64 PRELIMINARY; PRT; 690 AA.  
ID Q9GR64  
AC Q9GR64;  
DT 01-MAR-2003 (TREMblrel. 23, Created)  
DT 01-MAR-2003 (TREMblrel. 23, last sequence update)  
DT 01-MAR-2004 (TREMblrel. 26, last annotation update)  
DS Quinoxinoprotein alcohol dehydrogenase Adh IIB.  
DE Name=qdaa;

OS Pseudomonas putida.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=303;

RA TOYAMA H., FUJII T., AOKI N., MATSUHITA K., ADACHI O.;  
RT "Molecular Cloning of Quinolomoprotein Alcohol Dehydrogenase, ADH IIB,  
RL from Pseudomonas putida HK5.";  
RL Biosci. Biotechnol. Biochem. 67:1397-1400(2003).  
DR EMBL; AB091400; BAC1559.1; -.  
DR PDB; 1KV9; X-ray; A=23-690.  
DR GO; GO:0005746; C:mitochondrial electron transport chain; IEA.  
DR GO; GO:0030288; C:periplasmic space (sensu Gram-negative Bact. .; IEA.  
DR GO; GO:0005489; F:electron transporter activity; IEA.  
DR GO; GO:0006118; F:electron transport; IEA.  
DR InterPro; IPR001479; Bac\_PQQ.  
DR InterPro; IPR003045; Cytochrome\_BS.  
DR InterPro; IPR009056; Cytochrome\_C.  
DR InterPro; IPR003088; Cyt\_C1.  
DR InterPro; IPR002372; PQQ\_repeat.  
DR InterPro; IPR011047; Quin\_alc\_DH\_like.  
DR Pfam; PF00334; Cytochrom\_C; 1.  
DR Pfam; PF01011; PQQ; 3.  
DR SMART; SM00564; PQQ; 5.  
DR PROSITE; PS00364; BACTERIAL\_PQQ\_2; 1.  
DR PROSITE; PS00130; CYTOCHROME\_C; UNKNOWN 1.  
SQ SEQUENCE 690 AA; 74969 MW; F5198701D0937613 CRC64;

Query Match 20.0%; Score 612.5; DB 2; Length 690;  
Best Local Similarity 29.7%; Pred. No. 1.5e-33;  
Matches 181; Conservative 79; Mismatches 237; Indels 113; Gaps 20;

7 LRTSAVLLITAF-AAFAQVPTTDBLLANPPAGININFRQCNENRHSPLTQTADNVG 65  
5 LKSLMLCLATLALLAGVDAARAEQAGELSLGRFYAEGRFPLQIDASNR 64  
66 QLOLVW-----ARGEKAGAVQVTPIHNGVMTLANPGVITQALDAQTDLWERRQLP 119  
65 SLGLAWMDLIDNTRGEEA-----TPLFHGVITVTSMSRVLAVDAASGEKLM---RYDP 116  
120 AVATLNAQGD---RKRGVALYGTSLYFSSMNDHLTALDMENQGVQVFDVERSGEDGLTS 175  
117 EVAKVKARITSCCAVNRGVALMGDKYVGTLDRLTALDAKTKALW-----SQ 165  
176 NTGP-----IVANGVIVAGSTQYSPYGCFIGSDSANGEEIWRNHFIPQGE 224  
166 QITDPAPFYSITGAPRVKGVKVIIGNGAEYVGRG-FVSAVDADTGKLMR---FTVPGD 222  
225 -----EGDETNGNDFEARNMTG---VMQITVDPVTNIVFVGSTGVGPASETQ 269  
223 PALPYEHPEIRERAKTWQGD-QYKLDGGGGVWDSMAYDPELDLIVGTGSGSPNNREV 280  
270 RGTGGGLTGTNTRFAVRPDTEIWRHQTLPRNDWDOCTEEMVA--NVDVQPSAENE 327  
281 RSGGGGNLVLSSILAIRPDTGLAWHYQVTPGDSMDFTATQOITLAEINIDGKP----- 335  
328 GLRAINPNATGERRVLTGAPCKTGTMMSFDAISGEFLWAKRDNTYNTMLASIDETGLVTV 387  
336 -----RKVLMAQF-KNGFFVYVLDRTNGKLI-----SAKRGKVTW 369  
388 NEDAVL-----KELDVEYD-----VCPFLGGRDSSAALNPDITGYFLPNNACYDIM 436  
370 AEKVDLATGRPVAPGVRKEPELIVWSPFGAHNMHSMSPNGTGLVIVPQEVGVYR 429  
437 AVDOESALDVNTS---ATAKLAPGEMNGRIDALDISTGRITLMAEERPAANYSPVLT 493  
430 NEGDEFTKRAFNTAGFADATDVPAVVSALLANDPVKQKAMVPPYTHNGGTJST 489  
494 AGGVFVNGGTRFYRFAISOETGETLWQARLATVATGQAISYEVDGQVYLAIGG-----G 548  
490 AGNLVFGTAAAGOMHAYSDKGAALWQFEAGSGIVAAPMFFELAGQVYVAINAGMGVAT 549

QY 549 LTVGTQLNAP 558  
DB 550 LTGGESEYNLP 559

RESULT 10  
OHED\_COMTE STANDARD; PRT; 708 AA.  
ID OHED\_COMTE  
AC Q4644;  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 01-OCT-2004 (Rel. 45, Last annotation update)  
DE Quinolomoprotein ethanol dehydrogenase type I precursor (EC 1.1.99.-)  
DE (OH-EH1).  
GN Name=ghedH;  
OS Comamonas testosteroni (Pseudomonas testosteroni).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Comamonadaceae; Comamonas.  
OX NCBI\_TaxID=285;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15667;  
RX MEDLINE=96184549; PubMed=8654419;  
RA Stoetvogel J., Kraayveld D.E., van Sluis C.A., Jongejan J.A.,  
De Vries S., Duine J.A.,  
RT "Characterization of the gene encoding quinolomoprotein ethanol  
RT dehydrogenase of Comamonas testosteroni.";  
RL Eur. J. Biochem. 235:690-698(1996).  
RN [2]  
RP SEQUENCE OF 32-54 AND 477-490.  
RC STRAIN=ATCC 15667;  
RX MEDLINE=95324580; PubMed=7601151;  
RA de Jong G.A.H., Geertlot A., Stoetvogel J., Jongejan J.A., De Vries S.,  
Duine J.A.;  
RT "Quinolomoprotein ethanol dehydrogenase from Comamonas testosteroni.  
RT Purification, characterization, and reconstruction of the apoenzyme  
RT with pyrroloquinoline quinone analogues.";  
RL Eur. J. Biochem. 230:899-905(1996).  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE=86242113; PubMed=3521592;  
RA Green B.W., van Kleef M.A., Duine J.A.;  
RT "Quinolomoprotein alcohol dehydrogenase apoenzyme from Pseudomonas  
RT testosteroni.";  
RL Biochem. J. 234:611-615(1986).  
RN [4]  
RP CRYSTALLIZATION.  
RX MEDLINE=21536088; PubMed=11679760; DOI=10.1107/S0907444901013002;  
RA Oubrie A., Huizinga E.G., Rozeboom H.J., Kalk K.H., de Jong G.A.H.,  
Duine J.A., Dijkstra B.W.;  
RT "Crystallization of quinolomoprotein alcohol dehydrogenase from  
RT Comamonas testosteroni: crystals with unique optical properties.";  
RL Acta Crystallogr. D 57:1732-1734(2001).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (1.44 ANGSTROMS).  
RX MEDLINE=21671354; PubMed=1174714; DOI=10.1074/jbc.M109403200;  
RA Oubrie A., Rozeboom H.J., Kalk K.H., Huizinga E.G., Dijkstra B.W.;  
RT "Crystal structure of quinolomoprotein alcohol dehydrogenase from  
RT Comamonas testosteroni: structural basis for substrate oxidation and  
RT electron transfer.";  
RL J. Biol. Chem. 277:3727-3732(2002).  
RN [6]  
RP FUNCTION: Catalyzes the dye-linked oxidation of primary alcohols  
RN to the corresponding aldehydes and the (subsequent) oxidation of  
RN the aldehydes to carboxylic acids.  
RN - COFACTOR: Binds 1 PQQ group, 1 calcium ion, 1 heme and 1 iron ion  
RN per subunit.  
RN - SUBUNIT: Monomer.  
RN - SUBCELLULAR LOCATION: Periplasmic (Potential).  
RN - INDUCTION: By ethanol and butanol.  
RN - SIMILARITY: Belongs to the bacterial PQQ dehydrogenase family.  
RN - SIMILARITY: Contains 1 cytochrome c domain.  
RN -----  
RN This SWISS-PROT entry is copyright. It is produced through a collaboration







RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF176640; MAG09249.1; .  
 DR HSSP: Q46444; 1XB0.  
 DR GO: GO:0005746; C:mitochondrial electron transport chain; IEA.  
 DR GO: GO:0030286; C:periplasmic space (sensu Gram-negative Bact. .; IEA.  
 DR GO: GO:0005489; F:electron transporter activity; IEA.  
 DR GO: GO:0006118; F:electron transport; IEA.  
 DR InterPro: IPR001479; Bac\_PQQ.  
 DR InterPro: IPR000345; CytC\_heme\_B8.  
 DR InterPro: IPR009056; Cytochrome\_C.  
 DR InterPro: IPR003088; Cyt\_C1.  
 DR InterPro: IPR002372; PQQ\_repeat.  
 DR InterPro: IPR011047; Quin\_alc\_DH\_like.  
 DR Pfam: PF000034; Cytochrom\_C\_1.  
 DR Pfam: PF01011; PQQ\_3.  
 DR SMART: SM00564; PQQ\_5.  
 DR PROSITE: PS00364; BACTERIAL\_PQQ\_2; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; UNKNOWN 1.  
 DR SEQUENCE 695 AA; 75842 MW; D29A698A642E1A83 CRC64;  
 Query Match 19.4%; Score 594; DB 2; Length 695;  
 Best Local Similarity 30.7%; Pred. No. 2,9e-32;  
 Matches 193; Conservative 87; Mismatches 235; Indels 114; Gaps 30;  
 QY 6 LKRTSAVLLTLTA--PAFAQVTPITDEL--ANPAGWINYGRNENYHSPITQITA 61  
 DB 4 LTQSRRLVCLTLTLVAAGAAKVDAAIRASBOOSEWLSHGRTYAEGRFPLQIDA 63  
 QY 62 DNVGQLQVMA-----RGMEAGAVQVTPMHDGVMYLANP-GVIAQLD-AQTGDLLWEHR 115  
 DB 64 GNVGKGLAVYLDLENNRGLEA-----TPVSDGVYASLSMSRVAAVDLRSKRLMQ-- 116  
 QY 116 RQLPAVATLNAQGR-----KRGVALYGTSLYFSSWDNHLIALDMETGVVPEV 164  
 DB 117 -----FPQVDRGHSRYTCCDAVNRVAVLMKGYVALDGRLLIADAKGRELMS- 167  
 QY 165 ERSGGEDLTSTNTGPIVANG-VIVAGSTQYSPYCGFISGHDSATGEELMRNHFIP-- 220  
 DB 168 EORTDPKAPYSITGAPFVVKGYIINGGALYGVRG-FESAYDAETGKAMRFYTVPGDP 226  
 QY 221 -QDGE-----EGDETGNDFEARMNG-----VMGQITYPDVTNLVYGSTGVPASETQR 270  
 DB 227 AQPTEHELAEAAKTKGD--QYWKGGGGTWDMSVAYDELDLTYIG-TGNGSPNNRRI 283  
 QY 271 GTPGG--TLVGTNTRFARPDTEIYVRHOTLPKDNWDOCTFEMVAVAVDVPSAEMEG 328  
 DB 284 RSPGGGDNLY-LSSILALRPDSGKLMHVTTEGTEWDFATQOITLAL-----ELDG 336  
 QY 329 LRALNPAAATGERRVLTGAPCKTGTMSPFASGEFLMARDTNYTMIASID-ETGL-VT 386  
 DB 337 -----KPKKVLMOAP-KNGFVYLDRAATGELLSAEKFGKVTMAEKVDATGRPYE 385  
 QY 387 VNEDAVLKELDVEYDVCPTEFLGDRWSSAALNPDGTGFLPUNACYDIAVDOE-----F 442  
 DB 386 VPSRREKEGVVWVW-----PSSFGAHNHSMSPNQTGLMIFPYG---ELPGYIRNGAIF 438  
 QY 443 SALDVNTSATAKLAPGFENMGRIIDALISTGRTLSAER-----PAANY--SEVLST 493  
 DB 439 KKIDGNTGT-----GFSPTHEIPRDVAGSALLAMD PVKOREAWEPHSPFVNGTGLST 492  
 QY 494 AGGVVNGGDRYFRALSOETGETLWQARLATVATGQATSYELDGVQYAI-----GAAG 548  
 DB 493 AGNLVVGGRDGLAHYADKGRRLMSFPAQGTIVAAPISFSDGEQVAVVNAQMGGAAP 552  
 QY 549 LTVGQLNAPLAEALDSTSVGNAIYFAL 577  
 DB 553 LIGG---DALAPGVRLS---RLVYFKL 575  
 RESULT 13  
 Q89GY2 PRELIMINARY; PRT; 602 AA.  
 AC Q89GY2;

DT 01-JUN-2003 (TREMBLrel. 24, Created)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE Methanol dehydrogenase large subunit-like protein.  
 GN Name=mxaf; OrderedLocNames=dlr6213;  
 OS Bradyrhizobium japonicum.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Bradyrhizobiaceae; Bradyrhizobium.  
 RX NCBI\_TaxID=375;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=USDA110;  
 RX MEDLINE=22464398; PubMed=12597275;  
 RA Kaneo T., Nakamura Y., Sato S., Minamisaawa K., Uchiimi T.,  
 RA Sasamoto S., Watanabe A., Ideasa K., Iriyuchi M., Kawashima K.,  
 RA Kohara M., Matsunoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,  
 RA Tabata S.;  
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
 RT Bradyrhizobium japonicum USDA110.";  
 RL DNA Res 9:189-197(2002).  
 DR EMBL: AF005957; BAC51478.1; .  
 DR HSSP: P38539; 1G72.  
 DR GO: GO:0006118; P:electron transport; IEA.  
 DR InterPro: IPR002372; PQQ\_repeat.  
 DR InterPro: IPR011047; Quin\_alc\_DH\_like.  
 DR Pfam: PF01011; PQQ\_3.  
 DR SMART: SM00564; PQQ; 6.  
 KW Complete proteome.  
 SQ SEQUENCE 602 AA; 64779 MW; 96026018D3F7EE9D CRC64;  
 Query Match 19.3%; Score 593.5; DB 2; Length 602;  
 Best Local Similarity 28.0%; Pred. No. 2,6e-32;  
 Matches 181; Conservative 110; Mismatches 230; Indels 125; Gaps 26;  
 QY 5 TLRTSAVLLTLTAPAFQVTPITDEL--ANPAGWINYGRNENYHSPITQITA 61  
 DB 8 TYLGSAAALAVGSASA-----NDELIRKXQNPK--DWMPAGDVAANTRYSKLQINA 57  
 QY 62 DNVGQLQVMA-----RGMEAGAVQVTPMHDGVMYLANP-GVIAQLD-AQTGDLLWE 113  
 DB 58 QNVGQLQVAMFSTGVLRGHEG-----PLIIGNNVTYTPFNKVAIYDLSNNKTYWK 112  
 QY 114 HR-RQLPAVATLNAQGRKRGVALYGTSLYFSSWDNHLIALDMETGVVPEVRSGEDG 172  
 DB 113 YEPKDDPVVIRVMCCDVTNRLSGDGKILHQADVTNLVALDATGVVAMSATNGDPSKG 172  
 QY 173 LFSNTGSIYANGYIVAGSTQYSPYCGFISGHDSATGEELMRNHFIPQPEE----- 225  
 DB 173 QGTISALVVDKRYLVGSGSEFG-VQCHVTAYDLKSGKQYWR-AFSEGPDDQIKVDPAK 230  
 QY 226 -----GDETGNDFEARMNG-----VMGQITYPDVTNLVYGSTGVPASETQR 270  
 DB 221 TTSLGKPVGADSLKTWGD--QWKGGGCTGWMGSDPALNLVYVGS--NSTWNPK 285  
 QY 271 GTPGGTLYGNTTRAVRPDTEIYVRHOTLPKDNWDOCTFEMVAVAVDVPSAEMEGRL 330  
 DB 286 QRPQGNKK-SMTIFARBADTGMAKVYVQMTPHDMDVDGVNEMILSQOQINGA----- 338  
 QY 331 AINPAAATGERRVLTGAPCKTGTMSPFASGEFLMAR-----DNYNMI----- 376  
 DB 339 -----RKLITHTD-RNGLGVTMDRESELIVAEKDPKYNWISGVDMKNKSPYG 387  
 QY 377 -----ASIDETGLVTVNEDAVLKELDVEYDVCPTEFLGDRWSSAALNPDGTGYFLPLN 429  
 DB 388 RPKYLDAASTDKAG-----EDHNK-----GIOPALGTQDQAPAYSPDTQIFYVPTN 436  
 QY 430 NACVDIMAVDSEFALDVNTSATAKLAP--GFENMRIDAIDSTGRTLSARPRANY 487  
 DB 437 HVCWDYEPFKVSYTAGQY-VGATLSYTPROGESHMGNFTIAMDKTIVWSNKEQTSVW 495  
 QY 488 SPVLSTAGGVFNGGDRYFRALSOETGETLWQARLATVATGQATSYELDGVQYAI--G 545  
 DB 496 SGALATAGGVVFYGTLEGYLKAVDKSGKELYKTPSGIIGVNTYENGSGKQVAVVAVLSG 555

QY 546 AGGLTYGTOLNAPL-----AAIDSTSVGNATYFALP 578  
DB 556 VGGMA-GIGLAGLTDPTAGLAGVGYALSYTALGCTLYTFSLP 600

## RESULT 14

Q6MTW7 PRELIMINARY; PRT; 588 AA.

AC 06MTW7;  
DT 05-JUL-2004 (T-EMBLrel. 27, Created)  
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)  
DE 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)  
GN Alcohol dehydrogenase (Acceptor).  
OS CRFNames=NGR00479;  
OS Rhizobium sp. (strain NGR234).  
OC plasmid megaplasmid 2.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
OX NCBI\_TaxID=394;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=NGR234;  
RA Broughton W.J., Perret X., Staehelin C., Schmitz R.A., Raasch C.,  
RA Liesegang H., Gottschalk G., Streit W.R., to the EMBL/GenBank/DBJ databases.  
RL SDBL: AY316747; AA087251.1; "  
DR EMBL; AY316747; AA087251.1; "  
DR Interpro; IPR01047; Quin\_alc\_DH\_like.  
DR Pfam; PF01011; PQQ; 2.  
DR SMART; SM00564; PQQ; 7.  
KM Plasmid.  
SQ SEQUENCE 588 AA; 63953 MW; 5A6007CC5B3F6329 CRC64;

Query Match 19.0%; Score 582; DB 2; Length 588;  
Best Local Similarity 27.2%; Pred. No. 1.6e-31;  
Matches 174; Conservative 109; Mismatches 229; Indels 128; Gaps 26;

QY 7 LRISAVALLTAPAPAOVTPITDELLAN--PRAGEWINGRNOENYRHSPLTQTADNV 64  
DB 4 LMTAVTLLSLAGLASLSEV---TEEDLAKDATTGVLVTMGMRGQRYSPLFTLNKTV 60  
QY 65 GOLQLVWA-----RGEAGAVQVTPMIDGVMTLANPGDVIQALDAQTDLIWEHR 116  
DB 61 KNLLPANGFSLGSKRGQGES-----QPIYDGVMTITASYSRLYALDIKTGKELWQFDA 115  
QY 117 QLPRAVATLNAQGRKRGVALYGTSLYFSSWDNHLIALDMETGOVVDVBERGSGEDGLTSN 176  
DB 116 RLPE-GILPCCDVINRGAATYGDNIYFGTLDARLVALNRKTGDVWNKKIANYKEGY-SY 173  
QY 177 TTGPVANGVIVAGSTCOYSPYGC--ISGHSATGEELM-----RNHFIPOGEG-- 226  
DB 174 TAAPLVNGLITLINS--GGEFGIVGEVQARDAKTGELVWTRPVEIGHGVSYGKXSSMT 231  
QY 227 ---DETGNDEARMMTG---VWGQITYPVTNLVFGSTGVP-----ASETQRT 273  
DB 232 GTLNATWPGDL---WKTGGCATWLGSGSYDADTDLTVLFGTGNPAWNSHLNAGKPVENGK 288  
QY 274 GGTLYGTNTPAVRPDGEIVWRHQTLPNDWDECFEMMVANVAVOPSAEMEGRLAIN 333  
DB 289 GDNLYAA-SRLGINPENGEIKMHFQITPRGMDPDVNE-----VVFISKDGTRF- 339  
QY 334 PNAATGERRVLTGAPCKTGTWMSFDAASGEFL-----WARDNTYNTMIASIDETG 383  
DB 340 ---ATAADR-----NGFFVYLNREDGKFAVAPVKNITWAK-----GIDETG 378  
QY 384 LVTWNEADVILKELDVEVD-----VCPFTLGRDSSAALNPDGIYELPLNNACYDI 435  
DB 379 RPIYAENRPDGPSSADGKGQGVSVSPSLGKMMMPAYISQKSLFVPSNE----- 433  
QY 436 MAVDOESALDVNTSATAX-----LAPGFEN-NGRIDALDISTGRTILWSAERP 483  
DB 434 -----WGDIDWNEPISYKKGAAYLGAGFTIKPLFEDYIGSLKALIDPNNGEIKMEYKNA 486

QY 484 AANYSPVLSTAGGVNENGTDRTYFRALISQETGTLMOARLATVATGQAISYELDVQYIA 543  
DB 487 APLMGVATAGGLVFEVTPGECEFFKALIDDETGEELMSFQSGGVQGPITWEDGQYVS 546  
QY 544 I--GAGGLT--YGNQUNAPLAEALDSTSVGNATYFALPQ 579  
DB 547 VISGGMGAVPLMGGE---YAKKVTYLNQGMWWTFRLR 582

## RESULT 15

AA087251 PRELIMINARY; PRT; 588 AA.

AC AA087251;  
DT 02-MAR-2004 (T-EMBLrel. 27, Created)  
DT 02-MAR-2004 (T-EMBLrel. 27, Last sequence update)  
DE 02-MAR-2004 (T-EMBLrel. 27, Last annotation update)  
GN Alcohol dehydrogenase (Acceptor).  
OS Rhizobium sp. (strain NGR234).  
OC plasmid megaplasmid 2.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
OX NCBI\_TaxID=394;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=NGR234;  
RA Broughton W.J., Perret X., Staehelin C., Schmitz R.A., Raasch C.,  
RA Liesegang H., Gottschalk G., Streit W.R.,  
RT "Comparative DNA analysis of two large contigs of the Rhizobium sp.  
RT NGR234 megaplasmid 2."  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY316747; AA087251.1; "  
KM Plasmid.  
SQ SEQUENCE 588 AA; 63953 MW; 5A6007CC5B3F6329 CRC64;

Query Match 19.0%; Score 582; DB 2; Length 588;  
Best Local Similarity 27.2%; Pred. No. 1.6e-31;  
Matches 174; Conservative 109; Mismatches 229; Indels 128; Gaps 26;

QY 7 LRISAVALLTAPAPAOVTPITDELLAN--PRAGEWINGRNOENYRHSPLTQTADNV 64  
DB 4 LMTAVTLLSLAGLASLSEV---TEEDLAKDATTGVLVTMGMRGQRYSPLFTLNKTV 60  
QY 65 GOLQLVWA-----RGEAGAVQVTPMIDGVMTLANPGDVIQALDAQTDLIWEHR 116  
DB 61 KNLLPANGFSLGSKRGQGES-----QPIYDGVMTITASYSRLYALDIKTGKELWQFDA 115  
QY 117 QLPRAVATLNAQGRKRGVALYGTSLYFSSWDNHLIALDMETGOVVDVBERGSGEDGLTSN 176  
DB 116 RLPE-GILPCCDVINRGAATYGDNIYFGTLDARLVALNRKTGDVWNKKIANYKEGY-SY 173  
QY 177 TTGPVANGVIVAGSTCOYSPYGC--ISGHSATGEELM-----RNHFIPOGEG-- 226  
DB 174 TAAPLVNGLITLINS--GGEFGIVGEVQARDAKTGELVWTRPVEIGHGVSYGKXSSMT 231  
QY 227 ---DETGNDEARMMTG---VWGQITYPVTNLVFGSTGVP-----ASETQRT 273  
DB 232 GTLNATWPGDL---WKTGGCATWLGSGSYDADTDLTVLFGTGNPAWNSHLNAGKPVENGK 288  
QY 274 GGTLYGTNTPAVRPDGEIVWRHQTLPNDWDECFEMMVANVAVOPSAEMEGRLAIN 333  
DB 289 GDNLYAA-SRLGINPENGEIKMHFQITPRGMDPDVNE-----VVFISKDGTRF- 339  
QY 334 PNAATGERRVLTGAPCKTGTWMSFDAASGEFL-----WARDNTYNTMIASIDETG 383  
DB 340 ---ATAADR-----NGFFVYLNREDGKFAVAPVKNITWAK-----GIDETG 378  
QY 384 LVTWNEADVILKELDVEVD-----VCPFTLGRDSSAALNPDGIYELPLNNACYDI 435  
DB 379 RPIYAENRPDGPSSADGKGQGVSVSPSLGKMMMPAYISQKSLFVPSNE----- 433  
QY 436 MAVDOESALDVNTSATAX-----LAPGFEN-NGRIDALDISTGRTILWSAERP 483

Thu Nov 18 06:35:34 2004

us-10-802-682-8.rup

**Page 11**

Db 434 -----WGDIDINNEPISYKKGAAHIGAGFTTKPFDYIGISLALIPNBSGEIMWEYKA 486  
 QY 484 AANISPVLSAGSVFVNGSTDRYFRALSOETGETIMQARLATAVGAISYELDVGYIA 543  
 Db 487 APLMGVMAAGGLVVGTFPEGEFALDELTEBEELMSQTSGSVAGQPIWEGDGEQYVS 546  
 QY 544 I--GAGGLT--YGTOLNAPLAEIISTSGNAYVFAIPO 579  
 Db 547 VISMGGAVPLMNGE---VAKKVNTILOGGMMVTFRLPK 582

Search completed: November 18, 2004, 05:19:39  
Job time : 209 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: November 18, 2004, 04:39:43 ; Search time 42 Seconds

(Without alignments)  
1326.417 Million cell updates/sec

Title: US-10-802-682-8

Perfect score: 3069

Sequence: 1 MNPTTLRTSAVLLTAPA.....AEADSTSVGNALYFPALPO 579

Scoring table: BLOSUM62

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: PIR 79:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	597	19.5	708	2	SS2317
2	568	18.5	601	2	E95863
3	552	18.0	623	2	B83599
4	551.5	18.0	738	2	S14270
5	535.5	17.4	626	2	JQ0706
6	511.5	16.7	742	2	A49340
7	491.5	16.0	742	2	JS0326
8	435.5	14.2	573	2	S68591
9	399	13.0	796	1	JV0107
10	399	13.0	796	2	H85495
11	399	13.0	796	2	H90644
12	398.5	13.0	796	2	AG0523
13	375.5	12.2	801	1	SO0943
14	374.5	12.2	809	2	B98314
15	374.5	12.2	809	2	A12968
16	351.5	11.5	808	1	QPKEX
17	342.5	11.2	778	2	G98221
18	342.5	11.2	778	2	A13064
19	341.5	11.1	803	2	F83360
20	335	10.9	809	2	A55547
21	328	9.1	639	2	JC4881
22	274	7.3	221	2	A41378
23	172.5	5.6	407	2	H69064
24	168	5.5	524	2	A82580
25	159.5	5.2	407	2	C82804
26	152.5	5.0	668	2	C75264
27	151.5	4.9	839	2	D97033
28	144.5	4.7	613	2	F69424
29	140.5	4.6	3029	2	S76109

30	139	4.5	827	2	F64512	hypothetical prote
31	136	4.4	392	2	F91050	probable dehydroge
32	136	4.4	392	2	B65895	probable dehydroge
33	135	4.4	603	2	F72237	conserved hypotnet
34	133	4.3	392	2	G65027	hypothetical prote
35	132	4.3	2535	2	AC0304	probable hemolysin
36	131.5	4.3	424	2	T29127	hypothetical prote
37	130.5	4.3	1197	2	D82696	hypothetical prote
38	129.5	4.2	386	2	A82284	conserved hypotnet
39	128.5	4.2	2554	2	A83528	extracellular seri
40	128	4.2	1939	2	A82018	hypothetical prote
41	127.5	4.2	393	2	AD0350	probable lipoprote
42	127	4.1	392	2	AG0821	probable lipoprote
43	126	4.1	856	2	T00349	Avicelase III - As
44	126	4.1	946	2	F84280	ATP-dependent heli
45	124	4.0	796	2	T39962	coatomer complex b

## ALIGNMENTS

RESULT 1	SS2317	quinohemoprotein ethanol dehydrogenase (EC 1.1.99.-) type 1 precursor - Comamonas testost
C/Species:	Comamonas testostestroni	
C/Date:	08-May-1995	#sequence_revision 21-Jul-1995 #text_change 09-Jul-2004
C/Accession:	S62366; S62373; S65908; S52317	
R/Stoortvogel, J.; Kraayveld, D.E.; van Sluys, C.A.; Jongejan, J.A.; de Vries, S.; Duine, Eur. J. Biochem. 235, 690-698, 1996		
A>Title:	Characterization of the gene encoding quinohemoprotein ethanol dehydrogenase of	
A/Reference number:	S62366; MUID:96184549; PMID:865419	
A/Accession:	S62366	
A/Status:	preliminary	
A/Molecule type:	DNA	
A/Residues:	1-708 <ST01>	
A/Cross-references:	UNIPROT:Q46444; EMBL:X81880; NID:g663195; PIDN:CA57464.1; PID:g6631;	
A/Accession:	S62373	
A/Status:	preliminary	
A/Molecule type:	protein	
A/Residues:	32-54 <ST02>	
Ride Jong, G.A.H.; Geertlof, A.; Stoortvogel, J.; Jongejan, J.A.; de Vries, S.; Duine, Eur. J. Biochem. 230, 899-905, 1995		
A>Title:	Quinohemoprotein ethanol dehydrogenase from Comamonas testostestroni. Purification	
A/Reference number:	S65908; MUID:95324580; PMID:7601151	
A/Accession:	S65908	
A/Molecule type:	protein	
A/Residues:	32-50, 'X', 52-54; 477-483, 'X', 485-490 <DEJ>	
A/Experimental source:	ATCC 15667	
C/Genetics:		
A:Gene:	qhdh	
C/Superfamily:	glucose dehydrogenase (pyrroloquinoline-quinone)	
C/Keywords:	oxidoreductase, quinoprotein	
F1-31/Domain:	signal sequence #status predicted <SIG>	
F132-708/Product:	quinohemoprotein ethanol dehydrogenase type 1 #status experimental <XAI	
Query Match	19.5%; Score 597; DB 2; Length 708;	
Best Local Similarity	28.9%; Pred. No. 1.2e-33;	
Matches	178; Conservative 91; Mismatches 251; Indels 96; Gaps 21;	
QY	16	LTAPAFQVTPIT-----DELLANP-AGEKINYNQENYRHSPLTQITDN 63
DB	23	LGSAALFAQTGPAAQAAAVQVVDGFIRANVARTPDMPTIGVVYAETRYSRLOINMAN 82
QY	64	VGQDLTVARQMEA-GAVQVTPMTHDGYVYLANPQDVQALDAQDGLIWEHRQLPAVA 122
DB	83	VKDLGLANSYNLESTRGVETATPVVDGIMTVSASMSVHAIDTITGNIN----- 132
QY	123	TLNAQGR-----KEGVAYGTSLYFSSWDNHLALDMETGVQVDFVRSGED 171
DB	133	TYDPQIDSTGFKGCQDVVNRGVALMKGVVVGAWDGLILADATGKEVHONTFFEOX 192
QY	172	GLTSNTTGPVANGVYAGST-CQYSPYGCIFSGHDSATGELMRNHFIP-----QPGEE 225

```

Db      193 GSLITGAPRVFKVIGKGALEYGRC-YITAVDAETGERKRWFSVPDPSKFEDE 251
Qy      226 ----GDETW---GNDEPARMTGVMGOITVYDPTNLVFGSGTGVBPASETOGTPOGTL 278
Db      252 SMRAARITMDPSGKMWEDAGGGTMDSPFDALINTMYGTGNGSPMSHKVRS PKGGDNL 311
Qy      279 GINTPRAVRPDTGEIYWRHQTLPDRDWDDECTFEMVNAVVDQPSALEMGLAALNPNAI 338
Db      312 YLASIALPDPTGKYKMHVQETPGDNWDYTSQTPMLADIKI-----AG 355
Qy      339 GERVLVTAAGCTKTMWSEFPAASGEFLMARDTNYTMNLASIDETGL-VTVNEDAVLKELD 397
Db      356 KPRKVLHAP-KNGFFPVLDRTNGKFLSAKNFVPMWASGYDGHGKPGICI---AARDGS 411
Qy      398 VERDVCTPLFGDMSALINPDGTGYFLPLNACVDIADVDFEALVYNTSA----- 452
Db      412 KPDAVAPGPGYAHNMHPMSFNPTGLVYPAQVNPANLMD-DKMS---FNQAGPGKQ 466
Qy      453 -----TAKL---AGFEMWGRIDALIDISTGRTLSAERPAANYSPVLSTAGVVFNG 502
Db      467 SGTGNITAKRFNAPRPSKPFGLLMDPYAKKAASVEHSPMNGGTLTIGNVVFQGT 526
Qy      503 TDRYFALSGEETLMQARLATVATGQALSYELDGVQYIAIGAG-GLTYGTQLNAPLAE 561
Db      527 ADGRLVAYHAATGEKMEAPGTGTGVVAAPSTVYVADGRQYVAVAGVGYYGL-----AAR 581
Qy      562 AIDSTSVGNAYVFAL 577
Db      582 ATERGPG-TVYTFVV 596

```

## RESULT 2

```

E95863
alcohol dehydrogenase (acceptor) (EC 1.1.99.8) large subunit [imported] - Sinorhizobium
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C:Accession: E95863
R:Finan, T.M.; Weidner, S.; Wong, K.; Bulhmerster, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb PSYMB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: E95863
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-601 <KIR>
A:Cross-references: UNIPROT:Q92WY9; GB:AL591985; PID:CA048573.1; PID:G15140045; GSPDB:Q
A:Experimental source: strain 1021, megaplasmid PSYMB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Aboja, P.; Ampe, F.; Barloy-Hubler,
P.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federpiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.T.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SMD20173
A:Genome: plasmid
C:Keywords: oxidoreductase

```

```

Query Match      18.5%; Score 568; DB 2; Length 601;
Best Local Similarity 27.7%; Pred. No. 1e-31;
Matches 173; Conservative 101; Mismatches 243; Indels 108; Gaps 24;

```

```

Qy      23 AOVTPITDEL--LANPAGEMVINGRQENYRHSPLTQTADNVGQLVMA-----RG 74
Db      17 AQAFAFANDELOKLIDDE--NQMALQOTGDAYNLRYSKLDQINKDVGKQLQVAMFSTGVLRG 75
Qy      75 MEGAVQVTEMHIDGVWYLANP--GDVQALD--AQOTDLIMEHR--RLCPVATLNAAGDRK 131
Db      76 HEG-----SPVAVGDMLEVHTFPFNPVYALDLSKDGQIWKVZPKQDPNVIPWCCDTVN 130

```

```

Qy      132 RGVLYGTSLYFSSNDHLLALDMETGOVVDYERGGEGGLTNTTGP-VANGVIAGS 191
Db      131 RGVAYADNKKLPFHQADTTVALDCKTGKVMASVKNQDATGENTATVWVKDILVIGS 190
Qy      192 TCQSPYGCFLSGHDSATGELRNHPI-----PQCEBDEFTWGD 233
Db      191 GGEFGVNG-HYTAVSMADGKVLWRGVSMGPDSDTLIDPEKTYHGKRVGKDSGLTWEED 249
Qy      234 FEARWMTG---VMQITVPPVTLVVFYSGTGVBPASTOGTGGTLXGNTTFAVRPDT 290
Db      250 ---QMKIGGTTWGWYSYDPEENLVYGTGNTPESTNWTQS--PDNRM-SYTFARDVDT 303
Qy      291 GEIYWRHQTLPDRDWDDECTFEMVNAVVDQPSALEMGLAALNPNAATGERVLVTAAPCK 350
Db      304 GMAKMLVQMPHDEMDVDGVVEMILTBQIDGK-----DRKILTFED-R 346
Qy      351 TGTWMSDPAASGEFL-----WARDNTNYTMNLASIDETGLVYNEDAVLKEIDVEY 400
Db      347 NGFGYIMDRVYTGELVAEKYDPTVNMATE---VVMDEKSDKYG---RPQVAQYSTEQ 398
Qy      401 D-----VCPFLGGRDMSAALNPDTGYFLPLNACVDIADVDFEALVYNTSA 452
Db      399 NGEDNTTTCVCPALGTRKQDQPAAYSPKTELFPVPTNHVCMDEPFRVSYTAGQPY-VGA 457
Qy      453 TAKIAPGFEN--MGRIDAIDISTGRTLSAERPAANYSPVLSTAGCVYVNGGTRIFRA 509
Db      458 TLSMYPPKDSHGNGNFIANDNKEGKIKWSLPBFFSVMSALATAGDVVFYGLGELKA 517
Qy      510 LSGETGELMQARLATVATGQALSYELDGVQYIAI--GAGGLTYGTQLNAPLAEIDS-- 565
Db      518 VDAATGKELRYFKPPSGVIGVNMVYAREGQYVAVASGVGMA-GIGLAAGLINPTREGIG 576
Qy      566 -----TSVGNAYVFALPQ 579
Db      577 AVGVASLSNYTALGGLTVFKLPE 601

```

## RESULT 3

```

B83399
guinoprotein alcohol dehydrogenase PA1982 [imported] - Pseudomonas aeruginosa (strain PAC
Q:Accession: B83399
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Miroguchi, S.D.; Warriner, P.; Hickey, M.J.; Brj
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Liu,
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathog
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: B83399
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-623 <STO>
A:Cross-references: UNIPROT:Q924U7; GB:AE004624; GB:AE004091; NID:G9947973; PID:AA0537(
A:Experimental source: strain PA01
C:Genetics:
A:Gene: exaH; PA1982

```

```

Query Match      18.0%; Score 552; DB 2; Length 623;
Best Local Similarity 27.2%; Pred. No. 1.4e-30;
Matches 180; Conservative 110; Mismatches 230; Indels 142; Gaps 28;

```

```

Qy      3 PTLTIRTSA-----AVLLTAPAFAPVPTIDELLAN--PPAGEMVINGRQENYRHS 55
Db      9 PAGLLRPSLHCLAFVALGSGAALAK--DVTWEDINDKXTGDDVLYQWGTGTAQWSP 66
Qy      56 LTOITADNVGQLVMA-----RQMEGAVQVTEMHIDGVWYLANPBDVQALDAQT 107
Db      67 LKQVADVAVFELTPAWSVSPGDEKQGOESQAI-----VBDGVLYVYASVSRFLPALDAKT 121
Qy      108 GDLLWEHRRLP-----SAVATLNAQGRKGVVALYGTSLYFSSNDHLLALDMETGOVVF 162
Db      122 GKRLMTVYHRLPDDIRRCQVYV-----RGAALYGGKVFPGTLDASVVALNKRTGKVVW 175

```



```

QY 163 DVERGSEDGLTNTTGPPIYANG-----VIVAGST-CQYSPYCCFISGHSATGEBELMN 216
DB 176 KKKFADHAGAYTMTGAPTIVKDGKTGKVLHSSSGDEFVGRPLA-RPPTGGEIWMR 234
QY 217 HFIPOPEEG-----DETGNDFEA-----RMWTG--VMQITYPVTN 253
DB 235 PFV--EEMKRLKSKSTVTGDKAPBSPPDRNSPTGKESMSHGGAPOASFDPAETN 292
QY 254 LVFVYSTGVPASGTGTGPGG-----TLGNTNRPAPVPTGELVWRHQTLPEDNMQ 307
DB 293 TIIVGAGNPGPMNTWARTAKGNPHDYSLY-TSGQGVDPSSGEVWKFYQHTPNDAMDF 351
QY 308 ECFPEMVAWVDPQSAEMGLAINPNATGERRVLTGAPCKTGTMTSADAAGE----- 363
DB 352 SGNNELV-----FDYKADGKIVKATAHADNKGFFVVDNRNGKLQNA 395
QY 364 -----FLMARDNYNTMNASID-ETGLVTNED-AVLKELDVEY-----DVCPTLGR 410
DB 396 FPFVDNITWA-----SHIDLKTRGPVERREGQRPPEBQKHGKAVESSPPLGKR 446
QY 411 DMSAALNPTGTYFLPANNACVDINAVDOEFALDVYNTSATKLAPEF-----ENM 463
DB 447 NMPPMVYSQDTGFYVPPANH-----WKEDYTBESVSYTKSAYLGMGFRIKMYDDHY 499
QY 464 GRIDALDISTRTIMSAPERANYSPLYSTAGGVFNCGTDRYFALSOETGELMQAL 523
DB 500 GSRANDPVSGKVMHEKHLPLMAGVLATAGNLVFTGTGDYFKAPDAKSGELMKPOT 559
QY 524 ATATGQALSYELDGVQY--IAIGAGLT--YG--TQNLAPAEALDSTVGNALYVA 576
DB 560 GSGIVSPPTWEDGGEYLGVTVGAGVPLMGDMADILTRPVAQ-----GGSFVWR 612
QY 577 LP 578
DB 613 LP 614

```

RESULT 4  
S14270  
alcohol dehydrogenase, membrane-bound (BC 1.1.1.-) 72k chain precursor - Acetobacter pol  
C/Species: Acetobacter polyoxogenes  
A/Variety: strain NB11028  
C/Date: 21-Nov-1993 #sequence\_revision 26-Jul-1996 #text\_change 09-Jul-2004  
A/Accession: S14270  
R/Tamaki, T.; Fukaya, M.; Takemura, H.; Tayama, K.; Okumura, H.; Kawamura, Y.; Nishiyama  
Biochim. Biophys. Acta 1088, 292-300, 1991  
A/Title: Cloning and sequencing of the gene cluster encoding two subunits of membrane-bc  
A/Reference number: S14270; MUID:91159482; PMID:2001402  
A/Accession: S14270  
A/Molecule type: DNA  
A/Residues: 1-738 <TAM>  
A/Cross-references: UNIPROT:P28036; GB:D00635; NID:g16185; PIDN:BAA00528.1; PID:g216186  
A/Experimental source: strain NB11028  
C/Complex: heterodimer of 72k and 44k chains  
C/Superfamily: glucose dehydrogenase (pyrroloquinoline-quinone)  
C/Keywords: alcohol metabolism; membrane protein; oxidoreductase  
F/1-35/Domain: signal sequence #status predicted <Sig>  
F/36-738/Product: alcohol dehydrogenase 72k chain #status predicted <MAT>

Query Match 18.0%; Score 551.5; DB 2; Length 738;  
Best Local Similarity 27.0%; Pred. No. 1.9e-30;  
Matches 164; Conservative 91; Mismatches 253; Indels 99; Gaps 17;

```

QY 5 TLRTSAVLLTAPAAFAQVTP--ITDELL--ANPAGWINTGRNQENYRHSPLTOI 59
DB 15 TLTAGTICALLISGATVWASADGCGATGEALIHADHGNMWTYGRITVSDQYSPLDGI 74
QY 60 TANNQOLQVY-----ARGMEAGVQVTPMTHDSQVWMLAPRGVYQLMDQOTDLWE 113
DB 75 NNSNVGNLKLAMVLDLPTNRGGEG---TFLVIDSVMTATTNWMSMKKAVDADATGKLMS 129
QY 114 HRRQLPA-VATLNAQDRKRKGVALYGTSLYFSSWMDNHLIALDMETGVQVVEFY----- 164

```

```

DB 130 YDPRVGNIAKGCDDITVNRGAAYWNGKYVPTGFDGLIADAKGKLVMSVNTI-PPEAE 189
QY 165 ---ERGSDEGLTNTTGPPIYANG-VIVAGSTCQYSPYCCFISGHSATGEBELMNHIF 220
DB 190 LKQKSYVVDG-----APRIAKGVITIGNGSSEFGANG-FVSAPADATGKYDNRFFITV 242
QY 221 QGEEGD-----ETWGNDFEARMTG-----VMQITYPVTNLYFYG 258
DB 243 NFKNEPDAASDVLMNKAYQWTS-----PTGAWTRQCGCGTWDSTIVYDPAVLVLG 295
QY 259 STGVGPASGTGTGPGGTLGNTNRPAPVPTGELVWRHQTLPEDNMQDETFEMVANY 316
DB 296 VNGSPNNYKTRSEKGNLFGSLVAKPBTGEYWHFQETPMQDQMFSDQITMLD 355
QY 319 DVQPSAEMGLRAINPNATGERRVLTGAPCKTGTMTSADAAGEFLMARDNYNTMNAS 378
DB 356 PL-----NGETRHVIVHARKNGPFYIDAKTGFTISGRVYVVMASG 398
QY 379 ID-ETGLVTNEDAVLKELDVEYDVCPTFLGGRMSSAALNPTGTYFLPANNACV--- 433
DB 399 LDPKTRPIYNDALYTLTGKEWYGPEDLGHNFPAAAFSPKTLGYVITPAQVVPFYTN 458
QY 434 -----DIMAYDOEFALDVYNTSATKLAPEFENMGRIIDAIDISTGRTMSAERPA 484
DB 459 QVCGFTPHPDNMNIGLDMNKVGIID-SPEAKQAVKDLKGVITVMDQKQAEARVPHKG 517
QY 485 ANYSVYSTAGGVFNCGTDRYFALSOETGELTQARLATVATGQALSYELDGVQYTA- 543
DB 518 FWNQGIATGDDLLFOGLANGFPAHYATNGSDLFHPAASGIIAPVYTIANGKQYAV 577
QY 544 -IGAGLT 549
DB 578 EVNGGCI 584

```

RESULT 5  
J00706  
alcohol dehydrogenase (acceptor) (BC 1.1.1.99.8) alpha chain precursor - Methyllobacterium  
N/Alternate names: methanol dehydrogenase 62k large chain  
C/Species: Methyllobacterium extorquens  
C/Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004  
A/Accession: J00706; S07908  
R/Anderson, D.J.; Morris, C.J.; Nunn, D.N.; Anthony, C.; Lidstrom, M.E.  
Gene 90, 173-176, 1990  
A/Title: Nucleotide sequence of the Methyllobacterium extorquens AM1 moxF and moxJ genes  
A/Reference number: J00706; MUID:90373742; PMID:2116368  
A/Accession: J00706  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-626 <AND>  
A/Cross-references: UNIPROT:P16027; GB:M31108; NID:g150017; PIDN:AAA25380.1; PID:g150018  
A/Experimental source: strain AM1  
R/Nunn, D.N.; Day, D.; Anthony, C.  
Biochem. J. 260, 857-862, 1989  
A/Title: The second subunit of methanol dehydrogenase of Methyllobacterium extorquens AM1.  
A/Reference number: S04644; MUID:89350892; PMID:2504152  
A/Accession: S07908  
A/Molecule type: protein  
A/Residues: 28-50, 'XX', 53 <NUN>  
A/Note: the source is designated as Methyllobacterium extorquens AM1  
C/Comment: This enzyme oxidizes methanol to formaldehyde.  
C/Genetics:  
A/Gene: moxF  
C/Keywords: alcohol metabolism; oxidoreductase  
F/1-27/Domain: signal sequence #status predicted <Sig>  
F/28-626/Product: alcohol dehydrogenase (acceptor) #status experimental <MAT>

Query Match 17.4%; Score 535.5; DB 2; Length 626;  
Best Local Similarity 26.7%; Pred. No. 2e-29;  
Matches 168; Conservative 106; Mismatches 264; Indels 91; Gaps 22;  
QY 10 SAAVLLTAPAAFAQVTPITDELLANPAGE-WINYGRNQENYRHSPLTOITADNVQOLQ 68



```

QY 111 WEHRROPA-VATLNAQDGRKRGVALTGTSTIYFSWNNHIALDMETGOVVEDERSG 169
Db 126 LMQVPRKPGINADKGCDDVTYVRGAGYWKGVFMWTFPGRLYADAKTKKYMVAWNTTLP 185
QY 170 EDGLTSNTTGP1----VANGVIYAGS-TCQYSPYGEFISGHSATGSELMRNP1POG 223
Db 186 DASLGKQSYTVDGAVRVAKGLVILGNGAEEGARG-FVSAFDEATGKTKRFPYTPNNK 244
QY 224 EEGD-----ETWGDPEARMWT-----GWNGQITYPDVNTLVYGSRGVPA 265
Db 245 NEPDHAASDNLNKAAYKTWGP--KGAWVRQGGGTVDLSLYDPSLITLAVANGSEFW 302
QY 266 SETQGTGFGTLYGNTIRFAVRPDTGEIYMRHQTLPDRNDQECTFEKMWAVNDVQPSAE 325
Db 303 NYKXSEBISGNLFLGSIYALKEPGEIYMHFOATPMDQWDTYSVOQJMTLDMPVK---- 358
QY 326 MEGIRALNPNAATGE-BRVLTGAPCKTGTMWSFDAASEFLMARDYNTWNIAISDE-TG 383
Db 359 -----GENRVIYHAP-KNGFPYVLUDKTEGSEFSGKKYVQNNANGDLPLTG 404
QY 384 LVTYNEDAVLKELDVEYDVCPTFLGGRDWSSAALNPDIGIYFLPLNNAICYD1----- 435
Db 405 RPMYTPDGLYTLNGKFYWGIPGLGHNFMMAYSFKHLYVIRAHQJPFQYKQVGSFK 464
QY 436 -----NAVQEPBALDVYNTSATAKLAPFENNGRIDALDITSGRTLSNAEPRANYS 488
Db 465 PHADSKNNGULMTKNGKLPDTPEARFAYTK---DLHGWLADPYKMEYVMKIDHKGPNWG 521
QY 489 PVLSTAGGVNVNGDTRFYRALISOETGETTLQARLATATVQAOLASYELEDVQYITA--IGA 546
Db 522 GILATGDBLLFQGLANGEFHAYDATINGSDLYLKFDAGOSILAPMPTYSVNGKQYAVAVEGW 581
QY 547 GGL 549
Db 582 GGI 584

```

```

RESULT 8
568591
methanol dehydrogenase (EC 1.1.1.244) heavy chain - Methylophilus methylotrophus (strain DSMZ 568591)
C/Species: Methylophilus methylotrophus
A/Variety: strain W3a1
C/Date: 06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C/Accession: S68591
R/X/a: Z.; Dai, W.; Zhang, Y.; White, S.A.; Boyd, G.D.; Matthews, F.S.
J. Mol. Biol. 259, 460-501, 1996
A>Title: Determination of the gene sequence and the three-dimensional structure at 2.4 Å resolution of methanol dehydrogenase from Methylophilus methylotrophus
A/Reference number: S68591; MUID:96256524; PMID:8676383
A/Accession: S68591
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-573 <Xta>
A/Cross-references: UNIPROT:P36539, EMBL:U41040, NID:g1127819, PIDD:AAA83765.1; PID:g1127819
A/Note: the authors did not translate the codons for residues 1-2
C/Keywords: oxidoreductase

Query Match          14.2%; Score 435.5; DB 2; Length 573;
Best Local Similarity 25.3%; Pred. No. 1.6e-22;
Matches 151; Conservative 97; Mismatches 255; Indels 95; Gaps 26;

QY 38 AGEWINTGRNQCENYRHSPLTQTITADNVGQDQLVMARGMEAGV---GVTPMHIDGVNYL- 93
      ||| ||| ||| ||| :||| :: ||| :||| :||| :|||
Db 13 AGAAPPIATGGYYISQHSNSPLAQINKSNKYNAKAAVS--FSTGVNLNGHGAPLVIGDMMYVH 70
      ||| ||| ||| ||| ||| :||| :||| :||| :|||
QY 94 -ANGADVTCALDAQTGLIWEHR-RQLPAVAITLAQGDKKRGVALYGTSLYFTSSWDNHLI 151
      ||| ||| ||| ||| ||| :||| :||| :||| :|||
Db 71 SAFFNNIVYALNLDPGKIWQHKKQDASTKAWVCDDVDRLGLAYGAQIVKKGANSHL 130
      ||| ||| ||| ||| ||| :||| :||| :||| :|||
QY 152 ALMDETQQVVDFVERSGEGDELTSNTGPIYANGVIYAGSPTCYSPYGC--FIIGHDSAT 209
      ||| ||| ||| ||| ||| :||| :||| :||| :|||
Db 131 ALDAKTIRIMWEVSCDPKVGSTL-TQAPFAKHDTVLWG--CSGAELYGAGAVNAFLPKT 187
      ||| ||| ||| ||| ||| :||| :||| :||| :|||
QY 210 GEELAKR-----NHFIPOGE-EGETWGNDEAFARMTG---VMGQITY 248

```

```

Db      188 GELKMRAPFATGSDSVRLAKDFNSANPHYQFGLGTKEGD---AMKIGGTTMGWYAY 244
QY      249 DPTNIVFYGSGTVGASSTORGTGG-----TLVGTNTRFAVRPDGTGEIWRHQTLPD 303
Db      245 DPKLNTFFYGSGBAPMNETWR--FGDNKWTMTIWRGL-----DTGAKKGVQKTPHD 296
QY      304 NMPQECFEMMVANVDVQPSAEMEGRLRNPAANGSERVLGAPCKGTMTSPDAASE 363
Db      297 EMDPAGNQWVLTD---QP-----VNGKTPFLSHDRNGILTYTLNRENEN 339
QY      364 FLWARDTN-YTNMIASID-ETGLVTNEDAVLKELDVE-YDVCPTEFGSDWSSAALNPD 420
Db      340 LIYAEKVDAVAVNFKYVDLKTG-TVPADDEFATRMDHKNTICPSAMGFHNGVSDYDPE 398
QY      421 TGLYFLPLNNACVDIYNAQDEFSALDVYNTSATKLAP-----PENNGRIDALISGR 475
Db      399 SRLTYGLNLHICMDNEFPMLPYRAGOFF-VGATLAMYRPENPGPYTKENGQIHAFLDTTGG 457
QY      476 TWSAERPANYSPIVSTAGVVFNGGIDTRYFRLASQETGTLTQCARLATVATGQALSYE 535
Db      458 AKMTKEKKEPAANGSLTYRGGLVWYATLDGYVKKALDNKQKELMFPKMPSSGIGSPMTYS 517
QY      536 LQDVQYIA--IGAG-----GLTYGQNLNAPLA-----BAIDSTVGNALYFAL 577
Db      518 FKKQKQIIGSMYGVGMPGVGLVF--DLTDSAGLGAGVAFRELQNTQWGGGLMFSL 573

```

RESULT 9

JV0107  
glucose dehydrogenase (pyrroloquinoline-quinone) (EC 1.1.99.17) - *Escherichia coli* (strain ATCC 8739)

C:Species: *Escherichia coli*  
C:Date: 30-Jun-1992 #sequence\_revision 12-Dec-1997 #text\_change 09-Jul-2004  
C:Accession: D64735 JV0107; A45997; S45201; I41228  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col  
.A. Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.  
A:Reference number: A64720; PMID:97426617; PMID:9278503  
A:Accession: D64735

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
A:Residues: 1-796 <BLAT>

A:Cross-references: UNIPROT:P15877; GB:AEO00122; GB:U00096; NID:91786315; PIDN:MAC73335.1  
A:Experimental source: strain K-12, substrain MG1655  
R:Cleton-Jansen, A.M.; Goosen, N.; Fayet, O.; van de Putte, P.  
J. Bacteriol. 172, 6308-6315, 1990

A:Title: Cloning, mapping, and sequencing of the gene encoding *Escherichia coli* quinoprot  
A:Reference number: JV0107; PMID:91035240; PMID:2228962

A:Accession: JV0107

A:Molecule type: DNA  
A:Residues: 1-581, 'L', 60-148, 'H', 150, 'KRCRHT', 157-192, 'K', 194-796 <CITE>  
A:Cross-references: GB:X51323; NID:941553; PIDN:CAA55706.1; PID:941554  
A:Experimental source: strain K12  
R:Yamada, M.; Sumi, K.; Matsushita, K.; Adachi, O.; Yamada, Y.  
J. Biol. Chem. 268, 12812-12817, 1993

A:Title: Topological analysis of quinoprotein glucose dehydrogenase in *Escherichia coli* t  
A:Reference number: A45997; PMID:93286127; PMID:8509415

A:Accession: A45997

A:Status: preliminary

A:Molecule type: protein  
A:Residues: 1-20 <YAM>

R:Fujita, N.  
submitted to the EMBL Data Library, January 1994

A:Reference number: S45181

A:Accession: S45201

A:Molecule type: DNA  
A:Residues: 1-796 <FUJ>

A:Cross-references: EMBL:D26562; NID:9473770; PIDN:BAO0580.1; PID:9473791  
A:Experimental source: strain K-12 substrain W3110  
R:Yamada, M.; Asaoka, S.; Saito, M.H.; Yamada, Y.  
J. Bacteriol. 175, 568-571, 1993

A:Title: Characterization of the *gld* gene from *Escherichia coli* K-12 W3110 and regulator

A:Reference number: I41226; MUID:93123180; PMID:8419307  
 A:Accession: I41228  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-148; 'H',150, 'KRCCHT',157-192, 'K',194-665, 'H',667-796 <RES>  
 A:Cross-references: GB:D12651; NID:g21655; PIDN:BA02174.1; PID:g21656  
 C:Genetics:  
 A:Gene: gcd  
 A:Map position: 3 min  
 C:Function:  
 A:Description: catalyzes the oxidation of D-glucose to D-gluconic acid by ubiquinone  
 C:Pathway: respiratory chain  
 C:Superfamily: glucose dehydrogenase (pyrroloquinoline-quinone)  
 C:Keyword: oxidoreductase; pyrroloquinoline quinone; respiratory chain; transmembrane F  
 F/11-37/Domain: transmembrane #status predicted <TM1>  
 F/41-59/Domain: transmembrane #status predicted <TM2>  
 F/63-81/Domain: transmembrane #status predicted <TM3>  
 F/96-110/Domain: transmembrane #status predicted <TM4>  
 F/120-140/Domain: transmembrane #status predicted <TM5>  
 F/93-95/Binding site: ubiquinone (Arg, Asp) #status predicted  
 F/466/Active site: Asp #status predicted

Query Match 13.0%; Score 399; DB 1; Length 796;  
 Best Local Similarity 24.7%; Pred. No. 8.7e-20;  
 Matches 172; Conservative 91; Mismatches 247; Indels 186; Gaps 31;  
 Db 150 LSAADATPAEAKISPVADQ-----DWPAYGRNQEGQRFSPKQINADVNHNLKEAWVR 201  
 Qy 16 LTAFAFAQ-VTPITDELANPPAGEWINGRQENYRHSPLTQTADNVGQLQVYA-- 72  
 Db 73 -----RGMKAGAV--QVTPEIHGVMYLANPGDVICALDAQGDLIMEHRRQLPAVATLN 125  
 Qy 202 TGDVQKPNDEPGEITNEVTPIKVGDTLYLCTAHQRLPALDAASGKEMHNPDLKTNESP- 260  
 Db 126 AQGDRKRGVALYGTSLYFS-----MDNHLIALDMETGV-----VF 162  
 Qy 261 -QHTCGVSYHKAKEATSPVWADCPRIILPVNDGRILAINAENGKLCETPANKGVL 319  
 Db 163 DVERG--SGEDGLTSNTTGIYVANGVIV-AGS--TCQYS--PYGCFISGHSAATGEBELMN 216  
 Qy 320 NLQSNMPTDKGLYEPTSPPIITDKTIVMAGSVTDNFTSTRETSGVIRGPDVNTGELLMA- 378  
 Db 217 HFIPO-----PGEDEFTWGNDFEARMWTGWVGQITTPYNTLVFYSTGVGPASETQ 269  
 Qy 379 -FDGKADPNALPDSDEHTFTFNS-----PNSWAPAAVYAKUDLVYLPWGVTTPDIWGG 430  
 Db 270 RGTPEGTLTGNTFRFAVRPDTGEIIVWRHQTLPDNDQDCTEFEMVYANDVQPSAEMGL 329  
 Qy 431 NRTPEQERYASSI-LALNATGKLAWSYQTVHHDLMDDLPAPPTLADITV----- 480  
 Db 330 RAINPNAATGERRVLTGAPCKTGTMSPDAASGEFL-----WADDTNY----- 372  
 Qy 481 -----NGQKVVIYAPAKTGNIFVLDRNGELVVPAPKEKVPQGAAGDVTTPTQPF 532  
 Db 373 -----TNMIASIDETGLVTNEDAVLKEL--DVEYD-----VCPTEFLGGRD 411  
 Qy 533 SELSFRTKDLGSDAMWG-ATMPDQVCRVMFHQMYEGIFTPESEGTLVFPGNLGMFE 591  
 Db 412 WSSAALNPDTGIYF-----LPINNACY-----DIVAVDOFSALDVTNTSATA 454  
 Qy 592 WGGISVDPNREVAIANPMALPFVSKLIPRGGNPMEQPKAKGTGTESGIQPOGVYGV 651  
 Db 455 KLAPGFENM-----GRIDALDISTGRTLSA-----BRPANYSPV----- 490  
 Qy 652 TLNPFSLPGLPCKOPAMGYISALDLKTNEVWVKKIGTPQDSMPFMPVPVFPNMGPM 711  
 Db 491 ----LSTAGGVFENGCT-DRYFRALSOETGETLMQALATVATGA--ISYEIDGVQYIA 543  
 Qy 712 LGGISTAGNVLFIAADADNYLRAVYNSNGEKLWQGLP--AGGQATPMTYEVNGKQYIV 769  
 Db 544 IGAGGL-TYGTOLNAPLAEAIDSTVGNATVVFALP 578  
 Qy 770 ISAGHGSFGTKM-----GDYIVAYALP 792

RESULT 10  
 H85495  
 glucose dehydrogenase [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
 C:Accession: H85495  
 R:Perma: N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimmlanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: H85495  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-796 <STO>  
 A:Cross-references: UNIROT:Q88946; GB:AE005174; NID:G12512839; PIDN:AAG54428.1; GSPDB:Q  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: gcd  
 C:Superfamily: glucose dehydrogenase (pyrroloquinoline-quinone)

Query Match 13.0%; Score 399; DB 2; Length 796;  
 Best Local Similarity 24.7%; Pred. No. 8.7e-20;  
 Matches 172; Conservative 91; Mismatches 247; Indels 186; Gaps 31;  
 Db 150 LSAADATPAEAKISPVADQ-----DWPAYGRNQEGQRFSPKQINADVNHNLKEAWVR 201  
 Qy 16 LTAFAFAQ-VTPITDELANPPAGEWINGRQENYRHSPLTQTADNVGQLQVYA-- 72  
 Db 73 -----RGMKAGAV--QVTPEIHGVMYLANPGDVICALDAQGDLIMEHRRQLPAVATLN 125  
 Qy 202 TGDVQKPNDEPGEITNEVTPIKVGDTLYLCTAHQRLPALDAASGKEMHNPDLKTNESP- 260  
 Db 126 AQGDRKRGVALYGTSLYFS-----MDNHLIALDMETGV-----VF 162  
 Qy 261 -QHTCGVSYHKAKEATSPVWADCPRIILPVNDGRILAINAENGKLCETPANKGVL 319  
 Db 163 DVERG--SGEDGLTSNTTGIYVANGVIV-AGS--TCQYS--PYGCFISGHSAATGEBELMN 216  
 Qy 320 NLQSNMPTDKGLYEPTSPPIITDKTIVMAGSVTDNFTSTRETSGVIRGPDVNTGELLMA- 378  
 Db 217 HFIPO-----PGEDEFTWGNDFEARMWTGWVGQITTPYNTLVFYSTGVGPASETQ 269  
 Qy 379 -FDGKADPNALPDSDEHTFTFNS-----PNSWAPAAVYAKUDLVYLPWGVTTPDIWGG 430  
 Db 270 RGTPEGTLTGNTFRFAVRPDTGEIIVWRHQTLPDNDQDCTEFEMVYANDVQPSAEMGL 329  
 Qy 431 NRTPEQERYASSI-LALNATGKLAWSYQTVHHDLMDDLPAPPTLADITV----- 480  
 Db 330 RAINPNAATGERRVLTGAPCKTGTMSPDAASGEFL-----WADDTNY----- 372  
 Qy 481 -----NGQKVVIYAPAKTGNIFVLDRNGELVVPAPKEKVPQGAAGDVTTPTQPF 532  
 Db 373 -----TNMIASIDETGLVTNEDAVLKEL--DVEYD-----VCPTEFLGGRD 411  
 Qy 533 SELSFRTKDLGSDAMWG-ATMPDQVCRVMFHQMYEGIFTPESEGTLVFPGNLGMFE 591  
 Db 412 WSSAALNPDTGIYF-----LPINNACY-----DIVAVDOFSALDVTNTSATA 454  
 Qy 592 WGGISVDPNREVAIANPMALPFVSKLIPRGGNPMEQPKAKGTGTESGIQPOGVYGV 651  
 Db 455 KLAPGFENM-----GRIDALDISTGRTLSA-----BRPANYSPV----- 490  
 Qy 652 TLNPFSLPGLPCKOPAMGYISALDLKTNEVWVKKIGTPQDSMPFMPVPVFPNMGPM 711  
 Db 491 ----LSTAGGVFENGCT-DRYFRALSOETGETLMQALATVATGA--ISYEIDGVQYIA 543  
 Qy 712 LGGISTAGNVLFIAADADNYLRAVYNSNGEKLWQGLP--AGGQATPMTYEVNGKQYIV 769  
 Db 544 IGAGGL-TYGTOLNAPLAEAIDSTVGNATVVFALP 578  
 Qy 770 ISAGHGSFGTKM-----GDYIVAYALP 792



Db 713 GGPSTAGNVLFIAATADNLYRAVNMNGEKLWQSLP--AGGQATPMTYEVNGKQYVI 770  
QY 545 GAGGL--TYGTQNLAPLAIDSTSGNATYVFPALP 578  
Db 771 SAGHGSGFRTKM-----GDYIVAVPALP 792

## RESULT 13

500943  
glucose dehydrogenase (pyrroloquinoline-quinone) (EC 1.1.99.17) - Acinetobacter calcoacces  
C/Species: Acinetobacter calcoaceticus  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: S00943  
R/Cleton-Jansen, A.M.; Goosen, N.; Ogde, G.; van de Putte, P.  
Nucleic Acids Res. 16, 6228, 1988  
A/Title: Nucleotide sequence of the gene coding for quinoprotein glucose dehydrogenase F  
A/Reference number: S00943; MIMD:88289368; PMID:3399393  
A/Accession: S00943  
A/Molecule type: DNA  
A/Residues: 1-801 <CLE>  
A/Cross-references: UNIPROT:P05465; EMBL:X07235; NID:G38711; PIDN:CAA30222.1; PID:G38712  
A/Experimental source: strain LMD 79.41  
C/Function:  
A/Description: catalyzes the oxidation of D-glucose to D-gluconic acid by ubiquinone  
A/Pathway: respiratory chain  
C/Superfamily: glucose dehydrogenase (pyrroloquinoline-quinone)  
C/Keywords: oxidoreductase; pyrroloquinoline quinone; respiratory chain; transmembrane F  
F/3-35/Domain: transmembrane #status predicted <TM1>  
F/3-57/Domain: transmembrane #status predicted <TM2>  
F/61-79/Domain: transmembrane #status predicted <TM3>  
F/94-108/Domain: transmembrane #status predicted <TM4>  
F/118-137/Domain: transmembrane #status predicted <TM5>  
F/91,93/Binding site: ubi quinone (Arg, Asp) #status predicted  
F/471/Active site: Asp #status predicted

Query Match 12.2%; Score 375.5; DB 1; Length 801;  
Best local Similarity 23.7%; Pred. No. 3.8e-18;  
Matches 175; Conservative 89; Mismatches 248; Indels 227; Gaps 38;

QY 6 LRTSAVLLTPAPFAQVTPITDELANP-----AGEINTEGRQENYRHP 55  
Db 121 LSTSLALAIYLVMSIFNDQELNGEIKTPQPTAGAVPGVARSMDPAVGRQAGVRSY 180  
QY 56 LTQTLADNVGLOLVNARGMEAGAV-----QTPMIDGVYIYANPDVIOAD 104  
Db 181 LKQINDONVDLKVAMT--LRTGDLKTDNSGETNQVPIKGNMFLCTHAGQOLAD 238  
QY 105 AQTGDLIWEHRROLPAVALTNAQDRKGVATY-----GTSL-----Y 142  
Db 239 PATGKEKRRPDKTKDSF--OHLTCRGVYVDANNTEFATLSQSKSSSTQCPKRV 296  
QY 143 FSSMDNLILADMETGVNFDVERSGEDGLTS-----NTGPIVANG--VI 187  
Db 297 VPVNDGLVAVNADTGACTDF--GQNGVNLQEFMPYAVGNGNPSFGVIGSTIV 352  
QY 188 VAGS--TCQYS---PYGCFISGHSATGEELM-----RNHFIPOGEGGETWGNDE 235  
Db 353 IAGSVTNDYNSKEPSG--VIRGYDVTGKLLMVFDTGADBN--AMPGE-----GTFE- 401  
QY 236 ARMTGTWGGITTPVNTLVFYSTGVGPASFTORCPGGTLYG-----NTTRA 285  
Db 402 VHSNPAWPLADAKLIV--YVPTGV-----GTP--DIWGGRTKEKRYANSMLA 450  
QY 286 VRPDTEIWRHQTLPRDNDQECTFEMVANYDVOPSAEMEGLRANP--NATGERR- 342  
Db 451 INNSTGLVNFQTHHDLMDVPSQPLADIKNAAGTVPAIYVLTGKNAFVLDREN 510  
QY 343 -----VLTGAPCKTGTMMS-----FDAASEFLMA----- 367  
Db 511 GQPIVPTKEKVPQTVKRGQTKGEFYSKTQPSDNLAPQDKLTQDKMGATMLDQMC 570  
QY 368 ---RDNTYNTMTASIDETGLVTVNBDVAVLKELDEVYDVCPTFLGDRDSSALNDP--- 420

Db 571 RVSEKRLNVDGIYTPSENGTL-----VFPNULGVFEWQSGSVNPDROV 614  
QY 421 -----TGIVF-----LPLN--NACYDINAVDOEFSALDY-----YNTSAVAKLAP--GF-----E 461  
Db 615 AVNMPILGPFPSRLIPADPNRAQPAKAGTEQGVQPNVGVYGVETSAFSLPGLPCKP 674  
QY 462 NMGRIDADISTGRTLMSA-----BRPA-----NYSPLSTAGGVVFNQGT 503  
Db 675 AMGVVAGVDLKTTHVVKRIGTIRDSLPNLFLQIPAVKIGVPGLGSGISTAGVMEVGAT 734  
QY 504 -DRYFRALSOETGTLWQARLATVATGCA--ISYELDGVQYIAGAGL--TYGTQNLAPL 559  
Db 735 QDNVLRAPNTNGKULKEALP--AGGQATPMTYEVNGKQYVIYAGHGSGFRTKM----- 788  
QY 560 AEALDSTSGNATYVFPALP 578  
Db 789 -----GDYIVAVPALP 798

## RESULT 14

898314  
probable quinate dehydrogenase [imported] - Agrobacterium tumefaciens (strain C58, Cereor  
C/Species: Agrobacterium tumefaciens  
C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C/Accession: 898314  
R/Goodner, B.; Hinkle, G.; Galtung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A/Reference number: A97359; MIMD:21608551; PMID:11743194  
A/Accession: 898314  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-809 <KIR>  
A/Cross-references: UNIPROT:Q8UML9; GB:AE007870; PIDN:AAK90036.1; PID:G15160013; GSPDB:GT  
A/Map position: linear chromosome  
C/Superfamily: glucose dehydrogenase (pyrroloquinoline-quinone)

Query Match 12.2%; Score 374.5; DB 2; Length 809;  
Best local Similarity 23.2%; Pred. No. 4.5e-18;  
Matches 168; Conservative 80; Mismatches 238; Indels 239; Gaps 31;

QY 19 PAFPAQVTPITDELANPAGEMINYGROENYRHSPLTQITADNVGLOLVN----- 71  
Db 158 PAALTPVKEPTEQ-----KMEHYGNTSGGSRFVALDEITRDNIKLEAVMTYHTGDT 210  
QY 72 --ARGMEAGAVQTPMTIHGVMIYANPDVIOALDAGT-----DLMEHRROL 118  
Db 211 PISPGANGAEQDQTPLOVGVDTFLCTPHNNVIALDADTGTEKKAKETNSKSSVMMRCGL 270  
QY 119 -----PAVALTNAQDRKGVATGTSLYESSMDNLIALDMET--- 157  
Db 271 AVFPAKAPLQPTAPSGSTVPAAVVAG-----ALQGRILMNTVIAELIALDADTGA 324  
QY 158 -----GQVFDVERSGEDGLTSNTTGPIVANGVIVAG-----STQCYSPYCGFISGHD 206  
Db 325 CPDFTNGRVDLTKIGMGDAPDQYVLTSAPTLACTTVVAGRIADNVQVMDPGVAFGFD 384  
QY 207 SATGEELMRNH-----IPQGEGETWGNDEFAEMTGWGQITVDPVNTLVF- 256  
Db 385 VVTGELRKA--FDGNGPDITCLPPGQTYRSTN-----VMSMAYDDPLNLFM 433  
QY 257 -YGSTGVGPASFTORCPGGTLYGNTNRFAVRPDTEIWRHQTLPRDNDQECTFEMMV 315  
Db 434 PVGSPVDLYGATG--TPLDHKYGA--SMLALDATTGKEKVVQTVHNDLMD----- 481  
QY 316 ANVDVQPSAEMEGLRANPNNAATGERVILGAPC-----KTGTMWSPDAASG-----EF 364  
Db 482 FDIWQPS-----FVDFPRADG-----TSVPLVGTGAGGOLVLDKATGQGLTYVEE 529  
QY 365 LWARDNTYN-----MTASIDETGLVTVN-----DAVLEKEL--DVEYD-- 401



*This Page Blank (uspto)*